

A high-magnification scanning electron micrograph showing a dense, branching network of green, spherical Aspergillus spores. The spores are arranged in long, thin chains that branch out, creating a complex, tree-like structure. The background is dark, making the bright green spores stand out.

nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

FIGHTING FUNGI

How sensing melanin in *Aspergillus* spores sparks antifungal immune responses

PAGES 319 & 382

CONDENSED-MATTER PHYSICS

HIGH-ORDER STATES

Topological insulators turn a corner

PAGES 318, 342 & 346

GENOMICS

CHILDHOOD CANCERS

Genetic differences of tumours in children and adults


PAGES 316, 321 & 371

AGRICULTURE

SUSTAINABLE FARMING

Ways to boost productivity of smallholdings

PAGE 363

 [NATURE.COM/NATURE](https://www.nature.com/nature)

15 March 2018

Vol. 555, No. 7696

THIS WEEK

EDITORIALS

DATA Funders must explain benefits of management plans **p.286**

WORLD VIEW Use ocean floats to track regional carbon emissions **p.287**



GRIPPING Meet the robotic arm with a soft handshake **p.289**

AI diagnostics need attention

Computer algorithms to detect disease show great promise, but they must be developed and applied with care.

One of the biggest — and most lucrative — applications of artificial intelligence (AI) is in health care. And the capacity of AI to diagnose or predict disease risk is developing rapidly. In recent weeks, researchers have unveiled AI models that scan retinal images to predict eye- and cardiovascular-disease risk, and that analyse mammograms to detect breast cancer. Some AI tools have already found their way into clinical practice.

AI diagnostics have the potential to improve the delivery and effectiveness of health care. Many are a triumph for science, representing years of improvements in computing power and the neural networks that underlie deep learning. In this form of AI, computers process hundreds of thousands of labelled disease images, until they can classify the images unaided. In reports, researchers conclude that an algorithm is successful if it can identify a particular condition from such images as effectively as can pathologists and radiologists.

But that alone does not mean the AI diagnostic is ready for the clinic. Many reports are best viewed as analogous to studies showing that a drug kills a pathogen in a Petri dish. Such studies are exciting, but scientific process demands that the methods and materials be described in detail, and that the study is replicated and the drug tested in a progression of studies culminating in large clinical trials. This does not seem to be happening enough in AI diagnostics. Many in the field complain that too many developers are not taking the studies far enough. They are not applying the evidence-based approaches that are established in mature fields, such as drug development.

Many reports of new AI diagnostic tools, for example, go no further than preprints or claims on websites. They haven't undergone peer review, and might never do so. That would verify key details: the underlying algorithm code, and analyses of, for example, the images on which the model is trained, the physicians with which it is compared, the features the neural network used to make decisions, and caveats.

These details matter. For instance, one investigation published last year found that an AI model detected breast cancer in mammograms better than did 11 pathologists who were allowed assessment times of about one minute per image. However, a pathologist given unlimited time performed as well as AI, and found difficult-to-detect cases more often than the computers (B. E. Bejnordi *et al.* *J. Am. Med. Assoc.* **318**, 2199–2210; 2017).

Some issues might not appear until the tool is applied. For example, a diagnostic algorithm might incorrectly associate images produced using a particular device with a disease — but only because, during the training process, the clinic using that device saw more people with the disease than did another clinic using a different device.

These problems can be overcome. One way is for doctors who deploy AI diagnostic tools in the clinic to track results and report them, so that retrospective studies expose any deficiencies. Better yet, such tools should be developed rigorously — trained on extensive data and validated in controlled studies that undergo peer review. This is slow

and difficult, in part because privacy concerns can make it hard for researchers to access the massive amounts of medical data needed. A News story on page 293 discusses one possible answer: researchers are building blockchain-based systems to encourage patients to securely share information. At present, human oversight will probably prevent weaknesses in AI diagnosis from being a matter of life or death. That is why regulatory bodies, such as the US Food and Drug Administration, allow doctors to pilot technologies classified as low risk.

“Many in the field complain that too many developers are not taking the studies far enough.”

But lack of rigour does carry immediate risks: the hype–fail cycle could discourage others from investing in similar techniques that might be better. Sometimes, in a competitive field such as AI, a well-publicized set of results can be enough to stop rivals from entering the same field.

Slow and careful research is a better approach. Backed by reliable data and robust methods, it may take longer, and will not churn out as many crowd-pleasing announcements. But it could prevent deaths and change lives. ■

Russian research

The sleeping bear of Russian science could finally wake — and China can show it how.

Vladimir Putin will hardly be remembered as a patron of science. Not for Putin the scientific philosophy of dialectical materialism that helped to drive research in the former Soviet Union and that remains influential among many of his contemporaries. His long rule over Russia, as both president and prime minister, shows that he is more inclined to line up with the nation's Orthodox Church. His 2016 choice of an ultra-conservative religious historian as science and education minister was no accident.

But Putin, who is expected to win another six years in power in the Russian presidential elections on 18 March, did not get where he is today without being able to play both sides. He acknowledges — and has often said — that Russia's poor research and development capacity is an obstacle to economic growth and prosperity. His clique of political cronies includes scientists and research administrators. And their lobbying has not been in vain. Russian science spending has palpably (if by no means fully) recovered in recent years from near-collapse in the 1990s.

Outsiders recognize this: international sanctions in response to Russia's occupation of the Crimea have spared East–West research collaboration. And Russia's demanding education system continues

to produce a supply of excellent students and scientific talent. Yet, as discussed in a News story on page 297, too many Russian labs produce too little. Why is Russian science unable to take full advantage of its resources?

Putin would never admit it, but China — the other great power in the East — helps to highlight where Russia is going wrong. China also has a state-dominated economy, yet one that manages to create favourable research incentives. China's state-funded science system has its own problems, but is increasingly based on merit and competition and attracts foreign talent. Lively academic exchange with the West adds constant stimulus. And oriented towards the global market, industrial research in China operates in accordance with global demands, quality standards and management practices.

Russia, where anti-Western sentiment prevails, follows a quite different path. Fixed-term academic employment of postdoctoral researchers, who produce the majority of research in most countries, including China, is virtually unknown in Russian universities and research institutes. Instead, most academic scientists enjoy permanent positions for decades and feel little pressure to perform. Only a small fraction of public research spending comes as grants allocated through competition, with the rest being simply handed out by officials. The Russian Academy of Sciences — the country's foremost basic-research organization — is struggling to get on its feet after years of unproductive wrangling over money, direction and leadership.

Russia also puts too much trust in top-down innovation by state-owned companies — in aerospace and energy, for example. But these have struggled to develop, let alone export, innovative goods and ideas.

Russia's international political isolation, inflicted by Putin's erratic course and exacerbated by nationalistic rhetoric, is another obstacle. A recent crackdown on 'undesired foreign agents', including science-funding charities, sends a hostile signal to the outside world. Cronyism

and corruption start at the very top and undermine trust in research (and business) opportunities.

Putin clearly understands this. He has promised to increase science budgets further and to tackle funding bottlenecks that hurt competitive science. And on the face of it, a new national science strategy he launched in 2016 looked positive.

Under that plan, government funding was supposed to focus on a set of societally pressing topics — including energy research, health, digitalization, and security — which many other industrialized countries have also prioritized. Underperforming institutes run by the Russian Academy of Sciences would be restructured, or closed, and funding decisions spread over more shoulders to eliminate wheeling and dealing. None of this has happened yet.

Russia must wise up. If it's serious about science, then the steps are simple. Most urgently, the scattering of scarce resources indiscriminately among many large research organizations must stop. Grant money should be targeted towards the best projects and research groups. That's a goal that requires transparency, fair competition and international expertise to review the research — all eminently possible. A competitive programme to encourage young researchers to run independent groups for up to five years was launched last year by the Russian Science Foundation, a government-run grant-giving agency, and is a first step.

The country must go further, and remove notorious bureaucratic hurdles to doing science, including obstructive customs rules and import restrictions on research equipment.

A stronger Russia relies on a strong research base. Russian scientists — and the watching world — are tired of empty words. Putin defines himself as a man of action. Let's see some. ■

of data in question. That makes cross-disciplinary common standards unlikely, so research agencies need to engage with different scientific communities to create formats that best serve specific disciplines. To avoid a hotchpotch of standards, formats and data protocols — undesirable in our increasingly global scientific enterprise — research agencies in all parts of the world must engage.

An initiative for voluntary international alignment of research data-management policies, launched in January by Science Europe and the Netherlands Organisation for Scientific Research, is an important step in that direction. And existing data stewardship in particle physics and genomics shows that internationally aligned data governance not only is perfectly doable, but also has a positive impact on collaborative research. NASA pioneered this approach, setting up a centre in the 1980s to specifically curate the data from the Infrared Astronomical Satellite.

The message must now be passed on to scientists who work in fields less familiar with big data. Many of these, at all career stages, are worryingly unprepared. A survey of European researchers last year revealed that many have never been asked to provide a data-management plan, and that most are unaware of policies and guidelines already in place to help them. Only one-quarter of respondents to the survey, carried out by the European Commission and the European Council of Doctoral Candidates and Junior Researchers, had actually written a data-management plan, with another quarter saying they didn't even know what such a plan might be. There is nothing to suggest Europe is unusual in this.

Funders and universities, then, must ensure that the rationale of data management, and the basic skills of exercising it properly, become part of postgraduate education everywhere. Training and support must go further and be offered at every career level.

The laudable move towards open science — under which data are shared — makes the need for good data management more pressing than ever: there's no point in sharing data if they aren't clean and annotated enough to be reused. If you haven't got a plan for your data, you need one now. ■

Making plans

They sound dull, but data-management plans are essential, and funders must explain why.

Data are the alpha and omega of scientific and social research. A versatile good, they exist both as raw material for producing knowledge and, when processed and interpreted with an expert eye, the end product of the exercise.

So it might sound like a truism that researchers should conscientiously handle, preserve and — where appropriate — share the data they generate and use. The problem is that this can be hard to do.

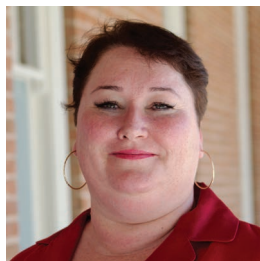
As science produces day by day a huge volume of data, it's a growing challenge to manage and store this information. To encourage this, many funders now ask applicants to submit a concise data-management plan with their grant proposals: effectively, a to-do list that details how an they plan to collect, clean, store and share the products of their research.

Such plans are important, and are something that *Nature* supports (we discuss them in detail in a Careers article on page 403). But to accelerate acceptance of what some might deem just another administrative burden, science funders and research institutions must work to streamline the process and to explain the need and benefits.

First, rigorously collected, well-preserved data sets — including meaningful descriptors or metadata — will help the data owners to reach solid, meaningful results. Second, they will help future investigators to make sense of and reuse data, thereby enhancing utility and reproducibility. Preserving comprehensive data, ideally for many years, also reduces the risk of duplicating science done by others.

Still, there is no single recipe for proper data management. The task varies according to the field of science, project size and the specific types

PAUL GOODMAN



Ocean sensors can track progress on climate goals

Uncertainties around carbon emissions will make climate agreements tough to enforce. The answer floats in the seas, says Joellen Russell.

Almost 200 nations have pledged to reduce their greenhouse-gas emissions under the Paris Agreement on climate change. We need a way to know whether they are succeeding.

For the accord to work, each country must track its net carbon output: the total carbon dioxide entering the atmosphere from fossil-fuel emissions and deforestation, minus that absorbed by growing plants. Both human emissions and land use are difficult to track precisely. Most nations extrapolate from government energy data, which are often incomplete, inaccurate or inconsistently reported.

As a result, the picture of who emits how much is murky. In 2015, for instance, China admitted that it had underestimated its annual coal consumption by up to 17%.

A solution lies in the seas. The ocean acts as a carbon sink, taking up about one-quarter of the world's global carbon emissions. If oceanographers can better monitor carbon uptake and then feed these data, along with other atmospheric and marine observations, into Earth-system models, we could track carbon output — including from land-use changes — regionally.

But we have too few measurements from the ocean. Fierce winds, high waves and long distances from port make working on ships expensive, inconvenient and dangerous. Vast swathes of sea need to be studied, especially in winter.

Enter the floating robots. Or rather, 1.3-metre tubes kitted out with batteries, sensors and a data-transmission system. A bladder lets the float sink to depths as great as 2,000 metres before surfacing to communicate its position and data to satellites. Many of these drifting floats are already active. Launched nearly two decades ago, the international Argo programme maintains an array of more than 3,800 floats to track the temperatures and salinities of oceans around the globe. All data are publicly accessible online.

But those measurements do not allow for assessments of marine carbon: researchers estimate this mainly from the few academic-research ships that sample at depth, and from the surface-only measurements they can get from commercial ships carrying sensors, as well as some sensors on deep-sea moorings, monthly campaigns off Hawaii and Bermuda, and ocean colour from satellite images. These sparse data, collated in the Surface Ocean CO₂ Atlas database, enable only rough, yearly estimates of carbon uptake.

To rectify this, my colleagues and I have been deploying Argo floats in the Southern Ocean that have been modified with biogeochemical sensors to measure oxygen and nitrate levels, pH and more. This project, called SOCCOM, has expanded our ability to measure carbon and carbon flux across seasons, at the ice edge, under the ice and in waters surrounded by ice. They work in storms, winds and heavy waves.

Other biogeochemical-sensor arrays also float in the North Atlantic, Mediterranean, North Pacific and Indian oceans. These are locally useful, but their data cannot be integrated across arrays. To help track regional carbon emissions, we need a global network of 'climate-quality' instruments: observations from each float must be sufficiently calibrated to permit comparisons to every other float, today and in the future.

Scientists from the main nations participating in Argo have agreed to the goal of adding biogeochemical sensors to some floats in the array, and the international plan for implementing a global network has been established.

What's stopping us? Mainly funding.

Building and maintaining the modified Argo network will require about US\$27 million annually. The US National Science Foundation has funded 200 modified floats, of which we have so far deployed 107. Ultimately, we need about 1,000 floats, each roughly 500 kilometres apart. A single modified float costs an estimated \$107,000, and will, over a 5- to 7-year deployment, collect thousands of measurements over a range of depths, down to 2,000 metres, and at the surface. For comparison, a day on a ship would cost at least \$50,000 and yield three depth profiles.

Globally flat funding for research — and the current political climate in the United States — makes it difficult to find extra funding to scale up even a highly successful, cost-effective project.

It is also technically challenging to build these floats, which are currently kitted out by hand to meet the standards necessary for calibrated measurements. Their manufacture must move from the laboratory to an industrial scale to produce the

number needed to build and maintain the array. Finally, to crunch so many data, we need to train more oceanographers.

Other ways of reducing uncertainty in the global carbon budget include NASA's Geostationary Carbon Observatory satellite, planned for the early 2020s, which will continuously measure greenhouse-gas emissions. Scientists have also proposed adding 'smart' nose cones to commercial airliners that could measure carbon dioxide and methane as the aeroplanes take off and land. These observations would nicely complement those from floating sensors.

Climate change and the grand experiment we are carrying out on our atmosphere make our future more uncertain than it has ever been. For the past 30 years, oceanographic research has been a priority of the US Navy, which believes that knowledge from the seas is essential to our collective interests and way of life. That is only more true today. ■

Joellen L. Russell is chair of integrative science at the University of Arizona in Tucson. She leads the modelling team of SOCCOM. e-mail: jrussell@email.arizona.edu

THE GRAND
EXPERIMENT
WE ARE CARRYING
OUT ON OUR
ATMOSPHERE
MAKES OUR FUTURE
MORE UNCERTAIN
THAN EVER.

SEVEN DAYS

The news in brief

PEOPLE

Genome pioneer

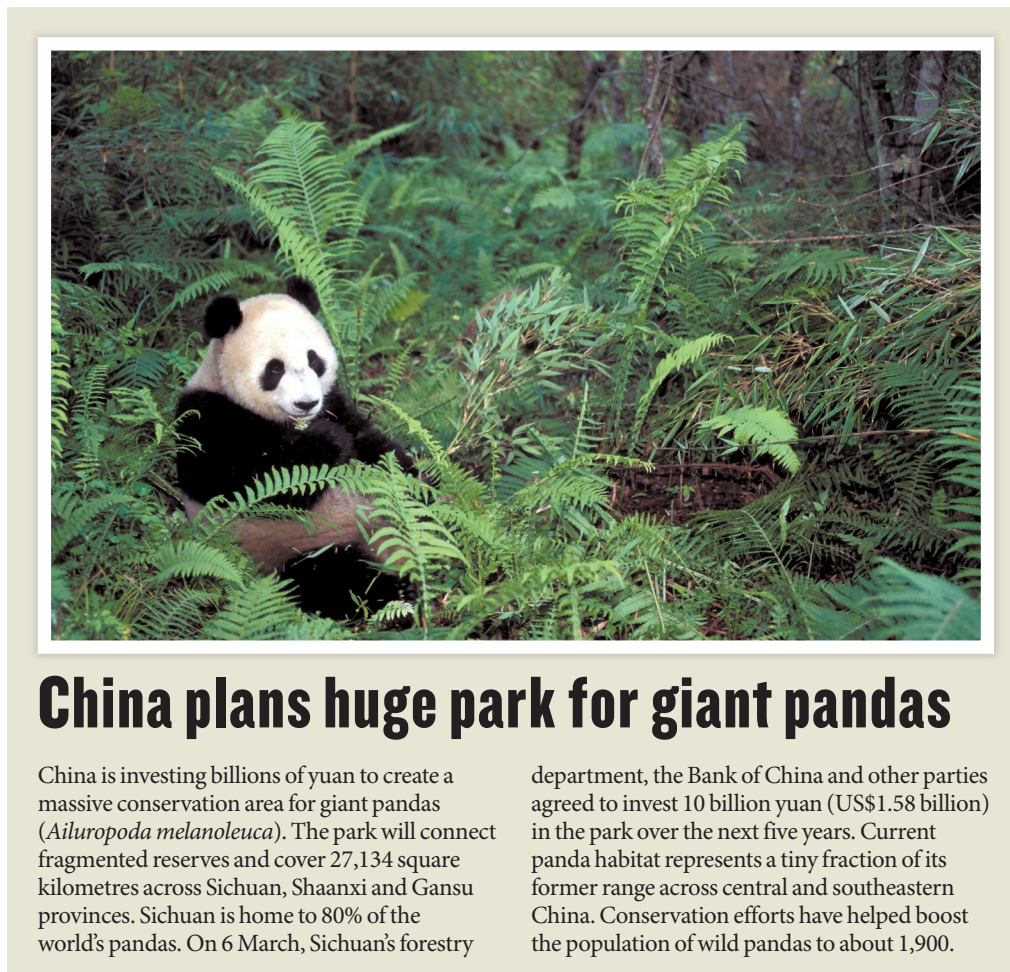
John Sulston, a Nobel-prizewinning British biologist, died on 6 March. Beginning in the 1970s at the Medical Research Council Laboratory of Molecular Biology in Cambridge, UK, Sulston painstakingly recorded each cell division in the development of the roundworm *Caenorhabditis elegans*, work that earned him a share of the 2002 Nobel Prize in Physiology or Medicine. Between 1992 and 2000, he served as the founding director of the Wellcome Sanger Institute in Hinxton, UK, where he played a part in the Human Genome Project.

Policy violations

On 7 March, Columbia University in New York City removed renowned neuroscientist Thomas Jessell. The move was based on the results of an investigation revealing “serious violations” of the school’s policies governing the behaviour of faculty in an academic environment, according to a university statement. It did not provide details of the violations in question. Jessell’s laboratory, which included around two dozen graduate students, postdoctoral researchers and other staff, will be shuttered. The university stated that it would help the students, researchers and staff to continue with their studies and careers.

NASA departure

NASA’s acting administrator, Robert Lightfoot, announced his retirement on 12 March — leaving the agency’s leadership even more uncertain as work begins on plans to return astronauts to the Moon. Last September, US President Donald Trump tapped Representative James



LYNN M. STONE/NATUREPL.COM

China plans huge park for giant pandas

China is investing billions of yuan to create a massive conservation area for giant pandas (*Ailuropoda melanoleuca*). The park will connect fragmented reserves and cover 27,134 square kilometres across Sichuan, Shaanxi and Gansu provinces. Sichuan is home to 80% of the world’s pandas. On 6 March, Sichuan’s forestry

department, the Bank of China and other parties agreed to invest 10 billion yuan (US\$1.58 billion) in the park over the next five years. Current panda habitat represents a tiny fraction of its former range across central and southeastern China. Conservation efforts have helped boost the population of wild pandas to about 1,900.

Bridenstine (Republican, Oklahoma) to head NASA. Lawmakers in the Senate have held up the nomination, citing Bridenstine’s political background and his statements expressing doubt about anthropogenic climate change. Lightfoot has run the agency in an acting capacity for 14 months.

POLITICS

China presidency

China’s Communist Party has abolished its constitutional two-term limit on the presidency. The rule change, approved on 11 March by the country’s parliament, allows current president Xi Jinping to remain in power for life.

Xi’s second term is due to end in 2023. The announcement prompted mixed reactions, with some commentators outside China critical of the plan. Chinese state media say the change is necessary to maintain the country’s stability. Xi has been a strong proponent of scientific research, and his government has enacted long-term plans to make the country a world leader in fields including artificial intelligence and quantum communication.

CONSERVATION

Big-game trophies

Hunters from the United States can now import trophies from

big game — such as elephants — from several African countries, the Associated Press reported on 6 March. The US Fish and Wildlife Service will consider imports of tusks, hides and other body parts from animals shot for sport, including African elephants (*Loxodonta africana*), lions (*Panthera leo*) and a type of antelope called a bontebok (*Damaliscus pygargus*). Decisions will be made “on a case-by-case basis”, according to a memorandum that the agency quietly released on 1 March. The decision overturns a ban on some big-game trophies put in place by former president Barack Obama’s administration.

AWARDS

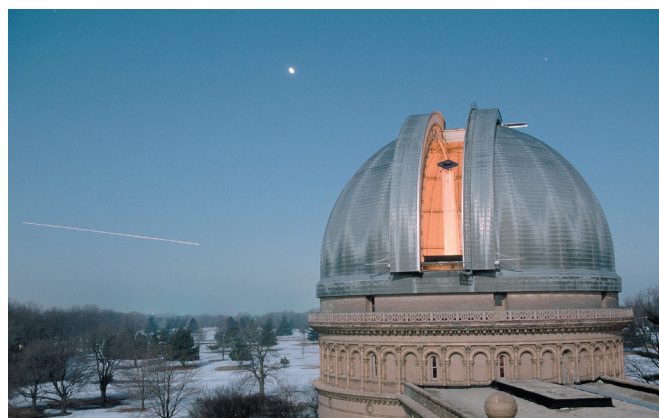
Brain prize

The 2018 Brain Prize has been awarded for work on the genetic and molecular basis of Alzheimer's disease. The Lundbeck Foundation in Copenhagen announced the winners on 6 March. Bart De Strooper at the University of Leuven, Belgium, Michel Goedert at the University of Cambridge, UK, Christian Haass at Ludwig Maximilian University of Munich, Germany, and John Hardy at University College London will share the €1-million (US\$1.23-million) prize. There is no treatment for Alzheimer's disease, which affects tens of millions of people worldwide.

EVENTS

Observatory closes

On 7 March, the University of Chicago in Illinois announced that it will cease operations at Yerkes Observatory, a storied astronomical institution in Williams Bay, Wisconsin, by 1 October. Telescopes at Yerkes, which opened in 1897, have been used to deduce the spiral nature of the Milky Way and to illuminate the behaviour of stars at the ends of their lives. But, in recent years, the university has shifted its research to newer, larger facilities and used Yerkes for



education and outreach. Those programmes will now move to its Hyde Park campus in Chicago.

UK strike hope

Union leaders and employers at the centre of a huge strike by UK academics have reached a deal that could see the walkout suspended. Academics at more than 60 institutions have been striking intermittently since 22 February over changes that would see their pension income go from having a guaranteed element to being entirely dependent on investment return, leaving them worse off in retirement. Universities UK, which represents the employers, said that the change was needed to tackle a deficit in the pension fund. On 12 March, the University and College Union (UCU) and Universities UK agreed on a

revised proposal that would return some defined benefits. As *Nature* went to press, UCU representatives were scheduled to discuss whether to approve the deal.

Climate review

The US National Academies of Science, Engineering, and Medicine (NAEM) released its review of a draft of the next US National Climate Assessment on 12 March. The assessment is a legally mandated report from government researchers on the state of climate-change science and is published every four years. Climate scientists and watchdog groups have been monitoring the process to ensure that officials in President Donald Trump's administration — some of whom have questioned climate science — do not interfere with the report. Thus far, the process has proceeded

smoothly. NAEM endorsed the draft assessment, as well as an accompanying draft report on the carbon cycle, suggesting only minor editorial changes.

BUSINESS

Cancer test

On 6 March, 23andMe, a DNA-sequencing company based in Mountain View, California, gained approval from the US Food and Drug Administration to offer genetic breast-cancer tests directly to consumers. The tests look for mutations in two genes — *BRCA1* and *BRCA2* — linked to breast and ovarian cancer, but they identify only three out of more than 1,000 known mutations in those genes. Other tests can identify many more mutations in the genes, but people need to see a physician to access the tests. The three mutations covered by the 23andMe product occur in about 2% of Ashkenazi Jewish women, but they are extremely rare in other populations.

POLICY

China reshuffle

China announced plans for a dramatic government overhaul on 13 March. Changes include plans for the Ministry of Science and Technology to oversee the National Natural Science Foundation of China, the country's main funder of scientific research, as well as another agency that recruits and certifies foreigners with expertise in areas including science, technology and economics. A restructured and expanded intellectual-property office will be given more power to enforce such rights, and a new ministry of ecological environment will be tasked with enhancing environmental protection and will be given significant powers to monitor pollution and enforce laws. It replaces the former Ministry of Environmental Protection.

TREND WATCH

An analysis in the United Kingdom shows that carbon dioxide emissions from fossil fuels fell by 2.6% in 2017, to levels last seen in 1890. The decrease was the result of a decline in the use of coal and natural gas, and it occurred despite a slight rise in the consumption of petroleum and oil. The analysis, by London-based climate watchdog Carbon Brief, is based on newly released government figures. The country's emissions have fallen every year since 2012, with the steepest drops in 2014 and 2016.

UK CARBON EMISSIONS HIT LOW

The United Kingdom's carbon dioxide emissions from fossil-fuel use have fallen steadily since 2012, with particularly large drops in 2014 and 2016.



NATURE.COM

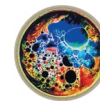
For daily news updates see:
www.nature.com/news

NEWS IN FOCUS

SEISMOLOGY Project off New Zealand probes quake risk **p.295**

BRAZIL Police action against researcher triggers protests **p.296**

RUSSIA Researchers worry about support for science **p.297**



CELL BIOLOGY To better understand the cell, look to the lava lamp. **p.300**

BSIP/UG/GETTY



Researchers are developing artificial-intelligence algorithms to detect breast cancer in mammograms.

MEDICAL RESEARCH

AI researchers embrace Bitcoin technology

Blockchain could let people retain control of data they contribute to health research.

BY AMY MAXMEN

Dexter Hadley thinks that artificial intelligence (AI) could do a much better job of detecting breast cancer than doctors do — if the screening algorithms could be trained on millions of mammograms. The problem is getting access to such massive quantities of data. Because of privacy laws in many countries, sensitive

medical information remains largely off-limits to researchers and technology companies.

So Hadley, a physician and computational biologist at the University of California, San Francisco, is trying a radical solution. He and his colleagues are building a system that allows people to share their medical data with researchers easily and securely — and retain control over the information. Their method, which is based on the blockchain technology

that underlies the cryptocurrency Bitcoin, will soon be put to the test. By May, Hadley and his colleagues will launch a study to train their AI algorithm to detect cancer, using mammograms that they hope to obtain from between 3 million and 5 million US women.

The team joins a growing number of academic scientists and start-ups that are using blockchain to make sharing medical scans, hospital records and genetic data ▶

► more attractive — and more efficient. Some projects will even pay people to use their information. The ultimate goal of many teams is to train AI algorithms on the data they solicit using the blockchain systems.

These efforts come as the public grows increasingly concerned about how tech giants mine and profit from personal data, including some medical information. In 2016, DeepMind, an AI company in London owned by Google's parent, Alphabet, became mired in controversy after press reports revealed that a branch of the UK National Health Service had given the company access to 1.6 million patient records without adequate consent. The information included names and sensitive information, such as whether a person had a sexually transmitted disease.

"Right now, Google and Facebook have siloed repositories of data about you that you have no control over," says Andrew Lippman, a computer scientist at the Massachusetts Institute of Technology in Cambridge. "But in the world of medicine, there is no Facebook." Using blockchain to secure and share decentralized medical information "could be a model of data-identity control" generally, he adds.

Blockchain is a distributed electronic system that records transactions in an expanding chain of 'blocks' that are extremely difficult to alter. To break into one block, a hacker would have to tamper independently with all the blocks that link to it — a daunting task.

In Hadley's study, blockchain will function as a series of switches that guide how data flow between participants, clinicians and researchers. Women taking part will be able

to give or revoke access to their data using an online portal, breastwecan.org, that relies on blockchain to secure data stored in the cloud.

The researchers plan to train their AI algorithm on millions of mammograms from healthy women and those with breast cancer. The goal is to classify tumours more precisely than doctors do; physicians miss up to one-quarter of cancers present in mammograms. The accuracy of an algorithm generally grows as it is trained on more, and more varied, data, just as a radiologist's ability to distinguish tumours improves with experience.

Hadley hopes that women will share their mammograms to improve breast-cancer screening generally — and to gain access to, and control over, information that has customarily been held by clinics. Women who participate in the study will be able to view their scans on breastwecan.org, along with standard clinical interpretations of their risk of breast cancer, based on tissue density, age and other known factors.

Other groups are developing blockchain-based marketplaces to broker data exchanges between individuals and companies or academic researchers — and arrange payment. One such effort is Nebula Genomics, a start-up co-founded by geneticist George Church of Harvard University in Cambridge, Massachusetts. Nebula aims to connect people who want their genomes sequenced with companies willing to pay for that service in return for access to the resulting data. People who pay for

their own sequencing will be able to sell access to their genetic information using Nebula; payment will come in the form of digital tokens that can be exchanged for US dollars.

Church says that Nebula will ensure that its partner companies keep any promises they make — on issues such as how long a company will retain a person's data. By contrast, when customers of genomic-sequencing firms such as 23andMe in Mountain View, California, consent to share their data for research, they largely relinquish control over how it is used. Many sequencing firms sell anonymized genetic data in bulk to biotechnology and pharmaceutical firms.

Giving people more control over their medical records could also yield more-immediate health benefits, Lippman says. He and his graduate students have developed a blockchain-based system for sharing health records, called MedRec, that will be tested at Beth Israel Deaconess Medical Center in Boston this year. The system allows users to insert information into their health records, including data from wearable electronic devices such as Fitbits. Clinicians and researchers could use these extra data, with permission, to tailor treatments.

Ultimately, Hadley says, the immense amount of routine medical data that physicians collect can yield medical advances only if the information is shared and studied. "We need to engage people so that they show us their data," he says. "So we need to think in medicine about the technologies that let us have good data governance, and blockchain happens to be one of them right now." ■ [SEE EDITORIAL P.285](#)

ENERGY

MIT renews push for fusion energy

Collaboration with company aims to feed grid in 15 years.

BY JEFF TOLLEFSON

The Massachusetts Institute of Technology (MIT) in Cambridge will work with a private firm to develop technology for producing energy from nuclear fusion within the next 15 years. If successful, the multimillion-dollar effort could help to unlock a virtually limitless source of pollution-free energy.

The approach — which has so far attracted US\$50 million — is based on high-temperature superconductors that have become commercially available in the past few years, the team announced on 8 March. The new generation

of superconductors will allow researchers from MIT and Commonwealth Fusion Systems (CFS) in Cambridge to strengthen the magnetic field that contains the hot-plasma fuel used in conventional tokamak reactors. That could pave the way for reactors that are smaller, cheaper and easier to build than those based on previous designs — including the international ITER project under development in southern France, which is over budget and behind schedule.

"It's about scale, and it's about speed," says Robert Mumgaard, chief executive of CFS. The company — an MIT spin-off — has attracted \$50 million from Italian energy giant ENI,

and plans to invest \$30 million of that sum in research and development at MIT over the next three years. Mumgaard says that the collaboration between academics and industrialists should help to drive fusion technology out of the lab and into the marketplace.

Fusing hydrogen atoms to form helium releases massive amounts of energy, which can be harnessed to produce carbon-free electricity. But sustaining the extreme temperatures that are required for this process in a confined space remains a daunting challenge that has defied most hopes and expectations to date.

CFS is the latest in a series of companies pursuing fusion energy as a clean-power source. Tokamak Energy, a company based near Oxford, UK, is also developing a tokamak reactor using high-temperature superconductors. But observers say that the MIT initiative is the most significant of its kind.

"There are no guarantees," says Stephen Dean, who heads Fusion Power Associates, an advocacy group in Gaithersburg, Maryland. But "if MIT can do what they are saying — and I have no reason to think that they

can't — this is a major step forward", he says.

The first challenge will be to transform a commercially available superconductor into a large, high-performance electromagnet, which could take around three years, says Martin Greenwald, deputy director of MIT's Plasma Science and Fusion Center. Within the next decade, the team hopes to develop a prototype reactor that can generate more energy

than it consumes. Then, they hope to develop a 200-megawatt pilot power plant that can export electricity to the grid.

Stewart Prager, former director of the Princeton Plasma Physics Laboratory in New Jersey, says it's good news that the MIT proposal is attracting private capital. But he warns that private investment won't be enough to make up for stagnant budgets in the US fusion

programme. "This funding for MIT is terrific, but there's just no way you are going to get the private sector to take on the full brunt of the fusion programme," Prager says.

For their part, MIT researchers hope that their work will generate more government interest in fusion research. "If we can change that narrative, we can potentially reinvigorate the rest of the programme," Greenwald says. ■

SEISMOLOGY

Drillers probe risk of big quakes in New Zealand

A major expedition is investigating the enigmatic sea-floor fault zone.

BY ALEXANDRA WITZE

An international team of geoscientists has launched a fully fledged onslaught to understand New Zealand's biggest earthquake and tsunami hazard.

On 11 March, the *JOIDES Resolution* drill ship began a two-month expedition to bore deep into the Hikurangi subduction zone off the east coast of New Zealand's North Island. There, the Pacific plate of Earth's crust dives, or subducts, beneath the Australian plate. The grinding of these geological titans has the potential to unleash a magnitude-9 earthquake and accompanying tsunami.

The drilling effort is part of a broader project to better understand the danger of the Hikurangi. "It's a major earthquake and tsunami hazard to the largest population centres, and it's not very well understood," says Laura Wallace, a geophysicist at the GNS Science research institute in Lower Hutt, New Zealand, and co-chief scientist on the upcoming cruise. The expedition will also give researchers the chance to probe the fault's role in a type of enigmatic slow-motion earthquake.

Whatever the scientists find will help to inform their understanding of seismic processes in other parts of the world with similar geologic settings, says Susan Schwartz, a geophysicist at the University of California, Santa Cruz.

Work kicked off in October, when researchers sprinkled nearly 300 seismometers in a dense array near the town of Gisborne on the North Island. Around the same time, two research vessels — the US *Marcus Langseth* and New Zealand's *Tangaroa* — deployed seismometers on the sea floor and blasted sound waves into the ocean crust to study its structure. Then, in December, the *JOIDES*



The Kaikoura earthquake in 2016 caused extensive damage on New Zealand's South Island.

Resolution did some initial drilling at three sites off the coast near Gisborne, to prepare for the bigger expedition that kicked off this week.

ANATOMY OF A DANGER ZONE

Together, the studies aim to build a detailed picture of the guts of the subduction zone. It is perhaps the largest geophysical-research effort in New Zealand's history, says Stuart Henrys, a geophysicist at GNS Science who led the deployment of the land seismometers. The governments of New Zealand, the United States, the United Kingdom and Japan are helping to fund research on the fault over five years.

One thrust of the work is to determine whether, and how often, the Hikurangi might

rupture in quakes as large as magnitude 8 or 9. A section of the fault offshore near Wellington is geologically locked and does not move, whereas a more northern part, near Gisborne, moves slowly. The seismic studies should help to illuminate the behaviour of rocks on either side of the fault and how that influences earthquake risk in both regions, Henrys says.

Another big question is the role of 'slow-slip' events akin to slow-motion earthquakes, in which the action unfolds over weeks or months, rather than seconds or minutes. Geologists aren't sure how slow-slip events influence the risk of larger quakes along a fault, but the Hikurangi is a natural laboratory for exploring that, Wallace says. Researchers ►

ANTHONY PHELPS/REUTERS

► can access it relatively easily because, off-shore of Gisborne, the subduction zone experiences the shallowest slow slip in the world, just a few kilometres below the sea floor.

DRILLING FOR ANSWERS

The Hikurangi usually sees slow-slip events once every year or two — including an episode triggered in November 2016 by the magnitude-7.8 Kaikoura quake on the South Island (L. M. Wallace *et al. Nature Geosci.* **10**, 765–770; 2017). “It basically lit up the subduction zone in slow slip,” says Wallace. What scientists learn about slow slip at the Hikurangi could

help them to better understand earthquakes in other slow-slip regions, including those off the coasts of Costa Rica, Mexico and Japan.

The *JOIDES Resolution* expedition aims to drill three holes into the area where the Pacific and Australian plates collide. This is likely to reveal what types of rock lie on either side of the Hikurangi fault, information that would enable researchers to better understand the physical properties of the place where earthquakes are generated.

One target is a thick layer of sediments covering the deep-diving Pacific crust. “Getting our hands on those sediments before they are

subducted will give us important insights into the frictional properties of rocks in the slow-slip zone,” Wallace says. Drillers will need to penetrate to 1.5 kilometres beneath the sea floor for scientists to truly understand this subducting crust and its role in Hikurangi quakes, says Nathan Bangs, a geophysicist at the University of Texas at Austin who led the *Langseth* cruises.

The drill team will install long-term observatories in two of the boreholes, roughly 400 metres beneath the sea floor, to monitor how pressure and temperature change during slow-slip events. ■

EVENTS

Backlash in Brazil against police probe of marijuana researcher

Investigation of a scientist in São Paulo sparks fear of restrictions to academic freedom.

BY CLAUDIO ANGELO

A police investigation targeting Brazil’s most prominent marijuana researcher has ignited a wave of protest among scientists. They say that the move by authorities from the state of São Paulo threatens research freedoms at a time when science in the country faces severe problems because of draconian budget cuts.

Police questioned Elisaldo Carlini, a retired professor of psychopharmacology at the Federal University of São Paulo (Unifesp), on 21 February on suspicion of inciting drug crime, according to authorities. They are still investigating the case and have not charged Carlini.

According to documents from Rosemary Porcelli da Silva, the public prosecutor in São Paulo state who requested opening the case against Carlini, she saw “in theory, strong hints of incitement” in a marijuana symposium that he had organized in May last year. One of the proposed guest speakers was the head of the Rastafari church in Brazil, who is still serving prison time under drug trafficking charges and did not participate in the symposium. Marijuana use, production and sale are illegal in Brazil. Da Silva declined to comment on the inquiry.

A PIONEER

Carlini, 87, is one of the pioneers of medical-marijuana research. He has investigated the drug since the 1950s, and has published several seminal papers on the anticonvulsive properties of cannabinoids. “Carlini

is an outstanding scientist,” says Raphael Mechoulam, a researcher at the Hebrew University of Jerusalem in Israel whose laboratory first isolated marijuana’s hallucinogenic compound, THC, in 1964.

“Nearly 40 years ago, his group and my group did the first clinical experiment with cannabidiol, a major cannabis compound, on epileptic patients,” Mechoulam says. A treatment that resulted from that work is used by people with epilepsy today.

“In more than 60 years of an academic career, I had never been questioned by law agents — until last month,” says Carlini. He says that last year’s meeting was scientific in nature and had nothing to do with inciting people to take drugs. “It’s a Kafkaian situation. I wonder what they think an old man can do with marijuana.”

On 1 March, researchers, students and staff at Unifesp gathered on campus to express their support for Carlini and to protest against what they perceived to be an attack on the university. As of 12 March, more than 50 scientific societies had signed a petition supporting the scientist. Another petition in defence of Carlini, organized by the Brazilian Society for the Advancement of Science (SBPC) and addressed to the São Paulo state authorities, had more than 34,000 signatures as of 12 March. Among the supporters is

former Brazilian president Fernando Henrique Cardoso, who called the inquiry into Carlini “an unacceptable coercion”.

ACADEMIC FREEDOM

SBPC president Ildeu Moreira says that the episode is a step backwards for academic freedom in Brazil, at a time when science there faces drastic funding declines.

Although it is not illegal to study cannabis in Brazil, current legislation makes it difficult. Research institutions cannot cultivate marijuana, and scientists must apply for a special government permit for any experiments with the drug or its components, which can delay their work. Brazil’s food and drug regulatory agency, Anvisa, is examining whether to authorize marijuana’s cultivation for research purposes.

Scientists say that they hope the Carlini case will highlight the difficulty that researchers in Brazil face when studying the medical uses of cannabis. Renato Filev, a neuroscientist at Unifesp, is trying to study whether cannabis can help people with alcoholism, and his animal experiments have shown encouraging results, he says. But clinical trials have been delayed because of the difficulties of getting a permit to bring the drug in from the Netherlands. Universities and ethics committees are afraid of the possibility of controversy or a police investigation, he says.

“I have fought for decades to show that marijuana is a serious plant,” says Carlini. “Dozens of countries have already regulated medical marijuana. The current legislation is a shame to Brazilian science and to Brazil.” ■

SCIENCE POLICY

Russian science chases escape from mediocrity

With Putin set to win another presidential term, researchers ask if science will be a priority.

BY QUIRIN SCHIERMEIER

After letting Russian science languish for years, Vladimir Putin has started to pay more attention. At a meeting of the Council for Science and Education last month, the Russian president promised that science and innovation are now top priorities. The presidential election on 18 March is likely to extend Putin's reign by another six years, but scientists inside and outside Russia wonder whether the country can reclaim its rich science legacy of Soviet times.

"Russia's research system isn't up-to-date any more," says polymer physicist Alexei Khokhlov of Lomonosov State University, a vice-president of the Russian Academy of Sciences. "It needs a thorough overhaul — otherwise the promises are just words."

Russia has a long way to go to recover its scientific might. Like many of the country's state institutions, its scientific infrastructure and workforce suffered after the break-up of the Soviet Union. Collapsing science budgets and scant salaries during the 1990s prompted thousands of Russian scientists to take up positions abroad, or to leave research altogether.

But there are signs that Russian science is starting to recover. Putin's government has gradually increased investments and public science spending over the past decade, and spending on research and development annually is now around 1% of gross domestic product (GDP) (see 'Russia rising').

SIGNS OF PROGRESS

The government earmarked 170 billion roubles (US\$3 billion) for fundamental research and development in 2018, a 25% rise over last year's basic science budget. The number of scientific papers produced in Russia more than



Vladimir Putin shakes hands with a robot built mainly with Russian-made parts.

doubled from 2006 to 2016, outpacing growth in both Brazil and South Korea. Russia is now in the top-ten countries in terms of number of research articles produced — ahead of Canada, Australia and Switzerland — according to statistics released in January by the US National Science Foundation.

"Russian science has greatly suffered, but we are now coming back to a reasonably predictable and well-organized situation," says Artem Oganov, a materials scientist formerly at the State University of New York at Stony Brook who took up a position at the Skolkovo Institute of Science and Technology in 2015. This private research university outside Moscow was created in 2011, in partnership with the Massachusetts Institute of Technology in Cambridge. "I would not have returned if there

had been no opportunity to do cutting-edge science here," says Oganov.

For all its progress, Russia's state-funded science still lags behind that of emerging science powers including China, India and South Korea, especially when it comes to translating discoveries into economic gains. Decades of underfunding, excessive state bureaucracy and entrenched opposition to reform within the country's sputtering research institutions are hampering competitiveness, says Khokhlov. "What we need are new ideas, new labs, fresh talent and more freedom and competition."

Many Russian researchers are vexed by state control of their work. An investigation by *Nature's* news team in 2015 found that many are obliged have their work vetted before they can submit it to foreign journals. ►



**MORE
ONLINE**

TOP NEWS



Millions of Chinese farmers reap benefits of crop experiment
go.nature.com/2iotkx1

MORE NEWS

- Neuron creation in brain's memory centre stops after childhood go.nature.com/2Inmoaz
- 'News' spreads faster when it's false go.nature.com/2fywfef
- What to expect from China's political meetings go.nature.com/2fw3ish

NATURE PODCAST



Geoengineering in Antarctica, and increasing the resolution of nuclear magnetic resonance nature.com/nature/podcast

► Researchers have also been aghast at a crackdown on science-funding charities deemed ‘undesirable’ foreign agents by the government, including the Dynasty Foundation and branches of the Open Society Foundations, founded by Hungarian-born US philanthropist George Soros.

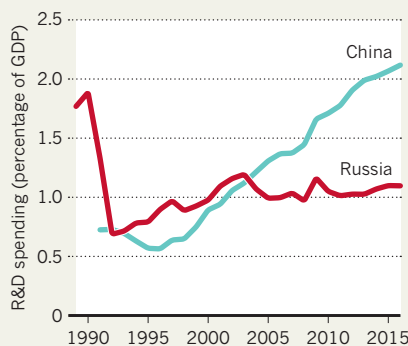
HOBBLED REFORM

Putin is eager to reduce Russia’s reliance on oil and gas exports. But research-driven efforts to diversify Russia’s economy, including a multibillion-rouble nanotechnology initiative launched in 2007, have not led to new blockbuster products or boosted the economy, say experts in Russian innovation. In 2016, the government launched a national science strategy that listed seven priority areas for state-funded research, including energy, health, agriculture and security. Scientist-led councils oversee funding and management of these efforts, a measure taken to cut down on cronyism by government officials and administrators.

Putin’s government also wants to push ahead with reform of the Russian Academy of Sciences, which operates more than 700 institutes in all fields of science. An evaluation completed in January found more than one-quarter of academy institutes to be ‘underperforming’ in terms of publications, research citations, patents and other metrics. These

RUSSIA RISING

As Russian science struggles to regain its prestige, emerging science powers such as China are pouring money into research and development.



institutes will now be asked to refocus under new leadership, or face closure, says Khokhlov.

The government plans to strengthen neglected university research, too. But, Khokhlov says, aspirations to bring at least 5 Russian universities into the global top 100 by 2020 seem to be unachievable because of scarce funding, poor infrastructure and the inability to attract talented scientists from abroad. Russian scientists will find “incomparably better” opportunities elsewhere, says Konstantin Severinov, a molecular biologist at

the Skolkovo Institute. “Money alone cannot build institutions.”

Long-simmering institutional problems are not the only drag on Russian science. Sanctions imposed in response to the annexation of Crimea in 2014 led to the suspension of civilian and military science and consultation under the NATO–Russia Council. Putin’s top science adviser, Andrei Fursenko, has been banned from entering the United States.

Russian support for Syria’s government in the country’s ongoing civil war, along with accusations of meddling in democratic elections, has soured relations with the West further. But, so far, geopolitics has not affected Russia’s participation in large international research efforts, such as the experimental fusion reactor ITER, under construction in southern France, or the European X-ray free-electron laser in Hamburg, Germany. Neither has it affected the country’s involvement in many smaller bilateral collaborations.

But Russian scientists do worry about the future. “Science doesn’t take place in a bubble,” says Fyodor Kondrashov, a Russian biologist working at the Institute of Science and Technology Austria in Klosterneuburg. “There are substantial barriers to doing competitive science in a politically isolated country. I don’t see how that should change as long as Putin holds the reins.” ■ SEE EDITORIAL P.285

SOURCE: OECD



Cell biology's new phase

Like oil in water, the contents of cells can separate into droplets. Finding out why is one of biology's hottest questions.

BY ELIE DOLGIN

When David Courson and Lindsay Moore arrived for a summer research placement in Woods Hole, Massachusetts, they expected to try some new techniques and play with high-end microscopes. As graduate students, they never imagined that they would help to solve a biological problem that had baffled researchers for more than 25 years.

Their instructors at the Marine Biological Laboratory asked them to decipher how pellets of RNA and protein called P granules form in worm embryos — a tall order given how long the structures had flummoxed biologists. Yet as soon as Courson and Moore started making movies of the process, they and their instructors could see something unusual happening under the microscope: the P granules were colliding and coalescing like blobs in a lava lamp.

Solid structures don't do that; only liquids

can. The P granules, they realized, were not hard kernels, as most researchers thought. Rather, they behaved like oil droplets in a bottle of vigorously shaken vinaigrette, first dispersing, then quickly fusing and blending into larger liquid blobs.

This process is a bread-and-butter concept in engineering, chemistry and physics, called liquid–liquid phase separation. It occurs whenever there's a force pushing two liquids apart, as when oil floats on top of water. Phase separation is common in nature and crucial in many industrial processes. Still, it wasn't an idea that Courson, a cell biologist now at Old Dominion University in Norfolk, Virginia, had come across. When he saw the P granules fuse like liquids, “it was a really neat moment”, he says, “but I didn't understand the scope or the scale of it”.

There was no more time to examine the process on the short summer course. But when the instructors, cell biologist Tony Hyman and his postdoc, biophysicist Cliff Brangwynne, returned to their lab at the Max Planck Institute of Molecular Cell Biology and Genetics (MPI-CBG) in Dresden, Germany, they ran some more experiments: they stuck worm gonads filled with P granules between two thin plates of glass and slid the plates past each other. Under the shear stress of the sliding plates, solids would smear out, but the granules merged, dripped and beaded up like rain drops on an umbrella.

That's when the magnitude of the discovery dawned on them. Phase separation might provide a way of concentrating certain molecules and excluding others to create order in the crowded chaos of the cell — an organizational feat that Hyman says biologists hadn't considered in any formal, quantitative way.

OIL ART: STEVE PAVLOVSKY/LIQUID LIGHT LAB

"It was just one of those questions people hadn't thought to ask," he says. Hyman and Brangwynne published their results¹ in 2009.

In the ensuing decade, scientists around the world have jumped on the idea that phase separation can explain how cells partition the molecules swarming inside them. These biological droplets could provide crucibles to speed up reactions, or quarantine unwanted or unneeded factors. "It's one of these in-hindsight, intuitive ideas. The second you hear it, it just makes a lot of sense," says Shana Elbaum-Garfinkle, a biophysicist at the City University of New York Advanced Science Research Center in New York City.

Not only is phase separation intuitive, but it seems to be everywhere. Droplets of proteins and RNAs are turning up in bacteria, fungi, plants and animals. Phase separation at the wrong place or time could create clogs or aggregate of molecules linked to neurodegenerative diseases, and poorly formed droplets could contribute to cancers and might help explain the ageing process (see 'Separate ways'). "It's a new paradigm that's really transforming our understanding of cell biology as a whole," says Elbaum-Garfinkle.

Yet some researchers think it's too early to say whether phase separation plays a major part in organizing the cell and causing disease. They suggest that it could simply be a side effect of chemical interactions, with little impact on cellular mechanics. Just because researchers can think of how a cell might use phase separation, it doesn't mean it's definitely happening, says Tim Mitchison, a cell biologist at Harvard Medical School in Boston, Massachusetts. "Those are just ideas right now. That's not really proof."

Researchers want that proof. "This is the multimillion-dollar question at this point," says Rohit Pappu, a computational biophysicist at Washington University in St. Louis, Missouri. "Is this some sort of by-product of sticky molecules being produced by the cell? Or did nature figure out how to use this advantageously?"

DROP BY DROP

As far back as 1899, US cell biologist Edmund Beecher Wilson anticipated that the main bulk of a cell, the cytoplasm, might include "a mixture of liquids" with "suspended drops ... of different chemical nature"². By the 1990s, researchers were beginning to speculate that phase separation might underlie disease or offer a general organizational principle in the cell.

These theories remained on the fringe, however. "It was mostly hypothetical," says Harry Walter, a retired chemical biologist who spent his career at the Veterans Affairs Medical Center in Long Beach, California. "It seemed logical that it should happen, but there was no scientific proof."

Some biologists had observed phase separation in specific, artificial circumstances — for example, while preparing proteins for X-ray crystallography studies. But few had

paid much attention to the phenomenon, or considered how it might relate to the formation of cellular compartments without borders.

Brangwynne and Hyman's 2009 report on worm P granules therefore came as a surprise — and initial reactions varied. Among worm biologists, "it ranged from those who thought it was total BS to those who thought that his group finally described the true nature of P granules", says Dustin Updike, who studies granule function at the MDI Biological Laboratory in Bar Harbor, Maine. And outside that research community, most scientists basically ignored

"The second you hear it, it just makes a lot of sense."

it. Fairly quickly, however, came solid evidence that phase separation in the cell was real.

In 2011, Hyman, Mitchison and Brangwynne — who set up his own lab at Princeton University in New Jersey that year — showed³ that the nucleolus, a dense cluster of genetic material and proteins in the cell nucleus, also exhibited droplet-like behaviours. A year later, independent groups led by structural biologist Michael Rosen and biochemist Steven McKnight, both at the University of Texas Southwestern Medical Center in Dallas, studied collections of proteins and RNA molecules in test tubes and found^{4,5} that the molecules were weakly attracted to each other, forming droplets and jelly-like blobs.

These 2012 studies, unlike Brangwynne and Hyman's earlier work, showed that phase separation could be reproduced in test tubes with fairly simple biochemical recipes. That made it a lot easier to study in the lab, Rosen says — and from there, "the field has exploded".

The boom began in early 2015, when a team led by Julie Forman-Kay, a structural biologist at the Hospital for Sick Children in Toronto, Canada, showed⁶ that a protein important for sperm function formed droplets in human cells. Before the year was up, more than half a dozen groups had published papers showing phase separation with their pet proteins. "We called it the flurry," says Elbaum-Garfinkle, who was a postdoc in Brangwynne's lab at the time, and lead author on one of the papers⁷.

Several of the proteins probed in the flurry were implicated in disease development. Researchers spotted it in motor neuron disease, or amyotrophic lateral sclerosis (ALS), a neurodegenerative condition characterized by abnormal clumps of protein in the nerve cells that control movement. Studies showed^{8,9} that the clumping process began when these proteins joined with other molecules, split from the surrounding cytoplasm and formed

droplets. These blobs grew increasingly gummy, ultimately turning rock-hard. "It's like taking room-temperature honey and putting it in the fridge," says Paul Taylor, a molecular neurogeneticist at St. Jude Children's Research Hospital in Memphis, Tennessee, who has documented phase separation in four proteins associated with the disease.

Those were some of the first concrete pieces of evidence that aberrant phase separation that turns liquids into solids might drive disease, says Jim Shorter, a protein biochemist at the University of Pennsylvania in Philadelphia. The process might be needed to partition cells, but when cells overdo it, he says, "they run the risk of forming structures that are perhaps more stable and solid and more difficult to reverse — and that's where you get into trouble".

SEPARATION ANXIETY

Several other diseases could be rooted in faulty phases. Just last month, Susanne Wegmann, a molecular biophysicist at Massachusetts General Hospital (MGH) in Charlestown, and her colleagues described¹⁰ phase separation in the tau protein, which aggregates into tangles in the brains of people with Alzheimer's disease. Phase separation "might be an initial trigger for aggregation", says Wegmann. This finding, she adds, "starts connecting the dots between these different neurodegenerative diseases".

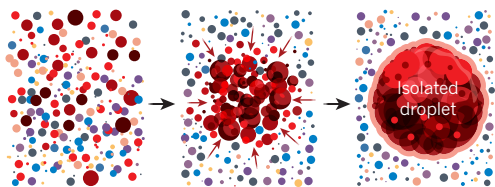
Errors in the phase-separation process could also prompt some cancers. Last year, a team led by MGH molecular pathologist Miguel Rivera identified¹¹ a protein implicated in Ewing's sarcoma that provokes activity in cancer-causing genes when it gathers near pieces of the genome linked to tumour formation. Aberrant phase separation allows the protein to build up in these areas. And last month, at the annual meeting of the Biophysical Society in San Francisco, California, structural biologist Tanja Mittag from St. Jude outlined how a protein that usually sequesters and destroys cancer-causing molecules inside droplets can instead provoke cancer when mutated, because the droplets no longer form.

These and other reported links to cancer and neurodegeneration prompted Hyman and Simon Alberti, a biochemist at MPI-CBG, to propose¹² that practically any ageing-associated disease could start when cells begin to lose control over phase separation. The body is in a constant struggle to keep its cellular house in order, "and at some point", Alberti says, "the system just breaks down".

But as well as damaging cells, phase separation can help them to adapt. Hyman and Alberti showed¹³ this year that when yeast cells are in stressful conditions of low pH, an evolved response triggers one of their essential proteins to form droplets to protect it. The gel disperses only when the pH rises and normal cellular functions can return. This finding dovetails with earlier work from Allan Drummond, a molecular and evolutionary biologist

Separate ways

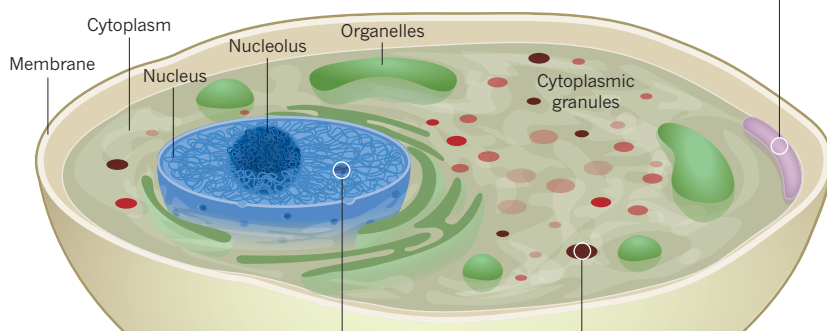
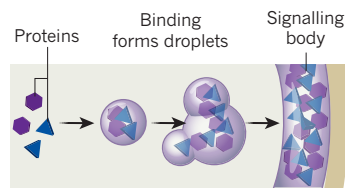
A cell's contents are thought to segregate through a process called phase separation to perform a wide variety of tasks. But flawed phase separation can also cause disease.



Physical forces between protein or RNA molecules can pull them apart or attract them to each other. Once the molecules reach a certain concentration, they can phase-separate, clustering similar components together to speed up reactions, or sequestering unwanted molecules.

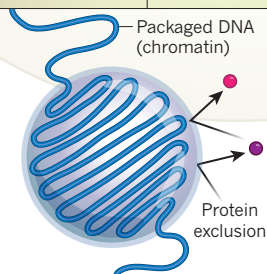
Signalling at the membrane

In neurons, proteins necessary for sending signals to neighbouring cells cluster at junctions and phase-separate to ensure smooth communication.



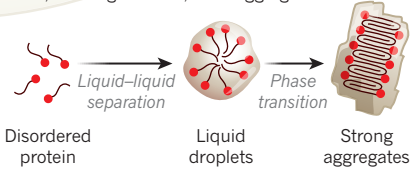
DNA packaging

In the cell nucleus, phase separation helps to compact unused DNA and quell its activity. Some proteins — possibly those involved in transcription — are excluded.



Drops become clogs

In amyotrophic lateral sclerosis, proteins that separate into liquid droplets can congeal over time, forming harmful, solid aggregates.



at the University of Chicago in Illinois, who reported¹⁴ on a different yeast protein that forms gels as a survival strategy at high temperatures.

Phase separation might thus be a general mechanism by which cells both sense stress and respond to it, says Drummond. “It’s like having the alarm also be the thing that turns on the fire hose,” he says.

In human cells, forming droplets could be more of an organizational strategy. Last year, biochemist Geeta Narlikar and her colleagues at the University of California, San Francisco, reported¹⁵ that phase separation helps to mothball parts of the human genome that are perpetually inactive and serve mainly a structural function. A team led by structural biologist Mingjie Zhang at the Hong Kong University of Science and Technology found¹⁶ that a piece of cellular machinery that helps brain cells receive signals is built using phase separation.

SHINING A LIGHT

Such studies are beginning to hint at some of the functions of liquid droplets in the cell, but they fail to explain why some components show phase separation, whereas others don’t. That frustrates researchers such as Hyman. “We have

to define the molecular grammar driving phase separation,” he says. And to do that, researchers needed a way of probing, controlling and bending the process to their own will in living cells. As Brangwynne puts it: “We needed tools.”

In a dark, windowless third-floor room in a 1970s concrete building at Princeton, Lian Zhu sits hunched by a microscope. A human cell speckled with red blobs lights up her computer screen, each dot denoting a throng of proteins that have phase-separated to form a nucleolus.

Zhu, a PhD student in Brangwynne’s lab, fires a blue laser at a spot in the cell, and within seconds new blobs emerge from the black ether. These are fluorescently tagged proteins from the nucleolus fused with a plant protein that, when illuminated with blue light, begins to cling to others of its type. Above a certain threshold, that triggers phase separation¹⁷.

This is what happens in Zhu’s cells. The red dots are droplets that appear and dance around the screen before starting to coalesce with others. “It’s like a magic trick,” Zhu says. By varying the dose of light, Brangwynne and his team can stiffen or loosen various liquid compartments inside living cells, triggering droplets to appear or disappear. Using the tool, Zhu has been

working to map the conditions under which nucleolus droplets form, showing how phase separation can occur in one part of the nucleus but fail to materialize in another.

Brangwynne hopes that the tool, dubbed optoDroplet, will bring new rigour to the study of phase separation. “We can now actually approach the level of detail that is standard for non-living materials, where you understand quantitatively what’s actually happening,” he says. That could be a huge boost for basic biological research, and could help researchers develop drugs by showing how much manipulation is needed to make or break droplets in cells.

Already, some companies are forming to pursue the idea of targeting phase separation to mitigate disease. Earlier this year, for example, a start-up founded by Ron Vale, a cell biologist at the University of California, San Francisco, received seed funding to search for drugs that break up RNA droplets associated with neurodegenerative conditions such as motor neuron disease and Huntington’s disease. Taylor is in discussions with investors about starting a company that will identify drug targets using an as-yet unpublished tool — Optogranule — that can recreate the pathology associated with phase separation in cells. The technique allows researchers to watch the neurodegenerative process happening in a dish in a matter of hours.

Others are taking a less guided approach to drug discovery. At MPI-CPG, for example, Hyman and Alberti have blindly screened a small library of approved drug compounds, looking for chemicals that put protein aggregates into a more fluid state. They have identified around 50 candidates. Now they are working out exactly how those drugs affect cellular function.

True progress in the field will require researchers to work out the rules governing how their drops and blobs form — and how to control them, says Brangwynne. “We need to take this to the next level.” ■

Elie Dolgin is a science journalist in Somerville, Massachusetts.

1. Brangwynne, C. P. *et al.* *Science* **324**, 1729–1732 (2009).
2. Wilson, E. B. *Science* **10**, 33–45 (1899).
3. Brangwynne, C. P., Mitchison, T. J. & Hyman, A. A. *Proc. Natl Acad. Sci. USA* **108**, 4334–4339 (2011).
4. Li, P. *et al.* *Nature* **483**, 336–340 (2012).
5. Kato, M. *et al.* *Cell* **149**, 753–767 (2012).
6. Nott, T. J. *et al.* *Mol. Cell* **57**, 936–947 (2015).
7. Elbaum-Garfinkle, S. *et al.* *Proc. Natl Acad. Sci. USA* **112**, 7189–7194 (2015).
8. Patel, A. *et al.* *Cell* **162**, 1066–1077 (2015).
9. Molliex, A. *et al.* *Cell* **163**, 123–133 (2015).
10. Wegmann, S. *et al.* *EMBO J.* e98049 (2018).
11. Boulay, G. *et al.* *Cell* **171**, 163–178.e19 (2017).
12. Alberti, S. & Hyman, A. A. *Bioessays* **38**, 959–968 (2016).
13. Franzmann, T. M. *et al.* *Science* **359**, eaao5654 (2018).
14. Riback, J. A. *et al.* *Cell* **168**, 1028–1040.e19 (2017).
15. Larson, A. G. *et al.* *Nature* **547**, 236–240 (2017).
16. Zeng, M. *et al.* *Cell* **166**, 1163–1175.e12 (2016).
17. Shin, Y. *et al.* *Cell* **168**, 159–171.e14 (2017).

COMMENT

NEUROSCIENCE Dispatch from the dark past of deep brain stimulation **p.306**

PALAEONTOLOGY Sex, power and DNA — tracking our ancestors around the world **p.307**

SUSTAINABILITY How can India manage its mountain of waste? **p.308**



ARCHAEOLOGY The Iceman cometh to a cinema near you **p.310**

JOE RAEDLE/GETTY



The village of Ilulissat in western Greenland is surrounded by icebergs that have calved from the Jakobshavn Glacier.

Geoengineer polar glaciers to slow sea-level rise

Stalling the fastest flows of ice into the oceans would buy us a few centuries to deal with climate change and protect coasts, argue **John C. Moore** and colleagues.

The ice sheets of Greenland and Antarctica will contribute more to sea-level rise this century than any other source. By mid-century, a 2°C increase is predicted¹ to swell the global oceans by around 20 centimetres, on average. By 2100, most large coastal cities will face sea levels that are more than a metre higher than currently.

If nothing is done, 0.5–5% of the world's population will be flooded each year after 2100 (ref. 2). For example, a 0.5-metre rise in Guangzhou, China, would displace more than 1 million people; a 2-metre rise would affect more than 2 million¹. Without coastal protection, the global cost of damages could

reach US\$50 trillion a year. Sea walls and flood defences cost tens of billions of dollars a year to construct and maintain².

At this price, geoengineering is competitive. For example, building an artificial island to host Hong Kong's international airport, which added 1% to the city's land area, cost more than \$20 billion. China's Three Gorges Dam, which spans the Yangtze River to control floods and generate power, is thought to have cost about \$33 billion.

We think that geoengineering of glaciers on a similar scale could delay much of Greenland and Antarctica's grounded ice from reaching the sea for centuries, buying time to address

global warming. In our view, this is plausible because about 90% of ice flowing to the sea from the Antarctic ice sheet^{3,4}, and about half of that lost from Greenland travels in narrow, fast ice streams. These streams measure tens of kilometres or less across. Fast glaciers slide on a film of water or wet sediment⁵. Stemming the largest flows would allow the ice sheets to thicken, slowing or even reversing their contribution to sea-level rise.

Geoengineering of glaciers has received little attention in journals. Most people assume that it is unfeasible and environmentally undesirable. We disagree. We understand the hesitancy to interfere with ►

► glaciers — as glaciologists, we know the pristine beauty of these places. But we have also stood on ice shelves that are now open ocean. If the world does nothing, ice sheets will keep shrinking and the losses will accelerate. Even if greenhouse-gas emissions are slashed, which looks unlikely, it would take decades for the climate to stabilize.

Is allowing a 'pristine' glacier to waste away worth forcing one million people from their homes? Ten million? One hundred million? Should we spend vast sums to wall off all the world's coasts, or can we address the problem at its source? Geoengineering is a political and societal choice, because people's reactions depend on how the issue is framed. Buttressing of glaciers needs a serious look. It should have fewer global environmental impacts than other proposals being discussed for reducing sea-level rise, such as injecting aerosols into the stratosphere to reflect sunlight and cool the planet.

To stimulate discussion, we explore three ways to delay the loss of ice sheets.

1. BLOCK WARM WATER

The Jakobshavn glacier in western Greenland is one of the fastest-moving ice masses on Earth. It contributes more to sea-level rise than any other glacier in the Northern Hemisphere. Ice loss from Jakobshavn explains around 4% of twentieth-century sea-level rise, or about 0.06 millimetres per year⁶.

Jakobshavn is retreating at its front. Relatively warm water from the Atlantic is flowing over a shallow sill (300 metres deep) and eating away at the glacier's base. Making the sill shallower would reduce the volume of warm water and slow the melting. More sea ice would form. Icebergs would lodge on the sill and prop up the glacier.

A 100-metre-high wall with sloping sides of 15–45° could be built across the 5-kilometre fjord in front of Jakobshavn glacier by dredging around 0.1 cubic kilometres of gravel and sand from Greenland's continental shelf (see 'Glacial geoengineering'). This artificial embankment, or berm, could be clad in concrete to stop it being eroded. The scale of the berm would be comparable with large civil-engineering projects. For example, ten times more material — 1 cubic kilometre — was excavated to build the Suez Canal. Hong Kong's airport required around 0.3 cubic kilometres of landfill. The Three Gorges Dam used 0.028 cubic kilometres of cast concrete.

Construction would be arduous and potentially hazardous in cold waters littered with icebergs. The reactions of local people would be mixed: although the project would create employment, large numbers of outside workers would have to be brought in. Ecology, fisheries and tourism could be affected. Glacier sediments supply nutrients for plankton growth, so marine ecosystems would be affected by increased turbulence during

construction of the berm and by the loss of sediment once the glacier was slowed.

Building such a berm would tell us whether glacial geoengineering is feasible, or if there would be unanticipated consequences. But the project would have only a small impact on 2100 global sea levels, given that Greenland's contribution is likely to be just 10–20 centimetres⁷. Antarctica will be the largest contributor, and geoengineering there will require larger and more challenging projects.

2. SUPPORT ICE SHELVES

Where Antarctica's ice sheets reach the sea, ice flows out as floating shelves. Pinned by rocks and islands, these platforms hold back the glaciers and limit how much ice reaches the sea. As the air and ocean around Antarctica warm, some ice shelves are becoming thinner, particularly those fringing the Amundsen Sea. In 2002, scientists were shocked at the collapse of 3,200 square kilometres of the Larsen B ice shelf, which is now only 30% of the size it was during the 1980s⁷. Half a dozen other shelves around the Antarctic Peninsula have shattered in the past 30 years.

Sheer cliffs are left behind when an ice sheet collapses. These crumble, accelerating the glacier's retreat⁸. The West Antarctic ice sheet is especially vulnerable because its bed rock lies below sea level and is deeper inland⁹. Warm ocean currents in the Amundsen Sea are melting the bottoms of floating parts of the glaciers, making the sheets more unstable.

The Pine Island³ and Thwaites⁴ glaciers in West Antarctica are the largest potential sources of sea-level rise over the next two centuries. Both glaciers are losing height and flowing more quickly than two decades

“Scientists were shocked at the collapse of the Larsen B ice shelf.”

ago. Pine Island Glacier reached a flow rate of about 4 kilometres per year in 2009, compared with 2.5 kilometres per year in 1996 (ref. 10). Models predict that, by 2150, these two glaciers might disgorge ice ten times faster than current rates, contributing 4 centimetres a year to global sea-level rise⁸.

One solution is to artificially pin the ice shelves in front of the two glaciers by constructing berms and islands, extended from outcrops or built on the sea floor. For example, the shelf buttressing Pine Island Glacier could be jammed by a berm located on Jenkins Ridge, a high point on the sea bed below the glacier. We estimate that this would require around 6 cubic kilometres of material, or 60 times more than would be needed to plug the Jakobshavn fjord. Relatively small artificial islands in other places — reaching up 300 metres from the sea bed — would require 0.1 cubic kilometres of material each. A large berm (10–50 cubic kilometres) in the open bay could prevent

warmer waters from entering.

Whether such engineering feats would successfully delay sea-level rise, and for how long, requires a better understanding of many factors. These include how the ocean circulates below ice shelves; how floating ice fractures and calves icebergs; and how glaciers slide and erode at their bases. A thorough study would be needed to determine the stresses that pinned ice shelves can sustain before they fracture. Models of ice dynamics should determine the most effective locations for pinning.

Material could be shipped to Antarctica from elsewhere in the world, or dredged or quarried locally. But it would be difficult in practice for engineers to work around the ice shelves, which grow and shrink as the glaciers, sheets and conditions fluctuate. Sea ice would also get in the way. Technologies might need to be developed to operate beneath floating ice. Major disturbances to local ecosystems would be expected and would require thorough assessment before and after pinning.

3. DRY SUBGLACIAL STREAMS

Fast-sliding ice streams supply 90% of ice entering the sea. As the ice slides over the glacier bed, frictional heat generates about 90% of the water at the base of the ice streams⁵. This water acts as a lubricant, speeding up the flow, which in turn generates more heat, and creates more water and slippage.

Glaciers in Greenland and at lower latitudes are relatively wet because their surfaces melt in summer, and rivers flow beneath them. In Antarctica, by contrast, there is little seasonal melting and much less water below the ice sheet. For example, the base of Pine Island Glacier releases about 50 cubic metres of water per second, which is only about 10 millimetres per year over the catchment area⁵. Removing this thin layer of water will slow the glacier, reducing frictional heating. The glacier will stall and its ice will thicken.

It is difficult to access the glacier's bed beneath one kilometre of ice, but there are precedents. The IceCube Neutrino Observatory at the South Pole has used jets of hot water to drill 60 holes to depths of 1,500–2,500 metres in the ice sheet. At Engabreen, Norway, a network of 5-metre-wide tunnels in the bedrock feeds 30–40 cubic metres of meltwater each second from the base of a glacier to the Svartsen hydropower plant. On the basis of current similar projects, we estimate that the cost of drilling the tunnels through rock beneath the Engabreen glacier was around \$500 million.

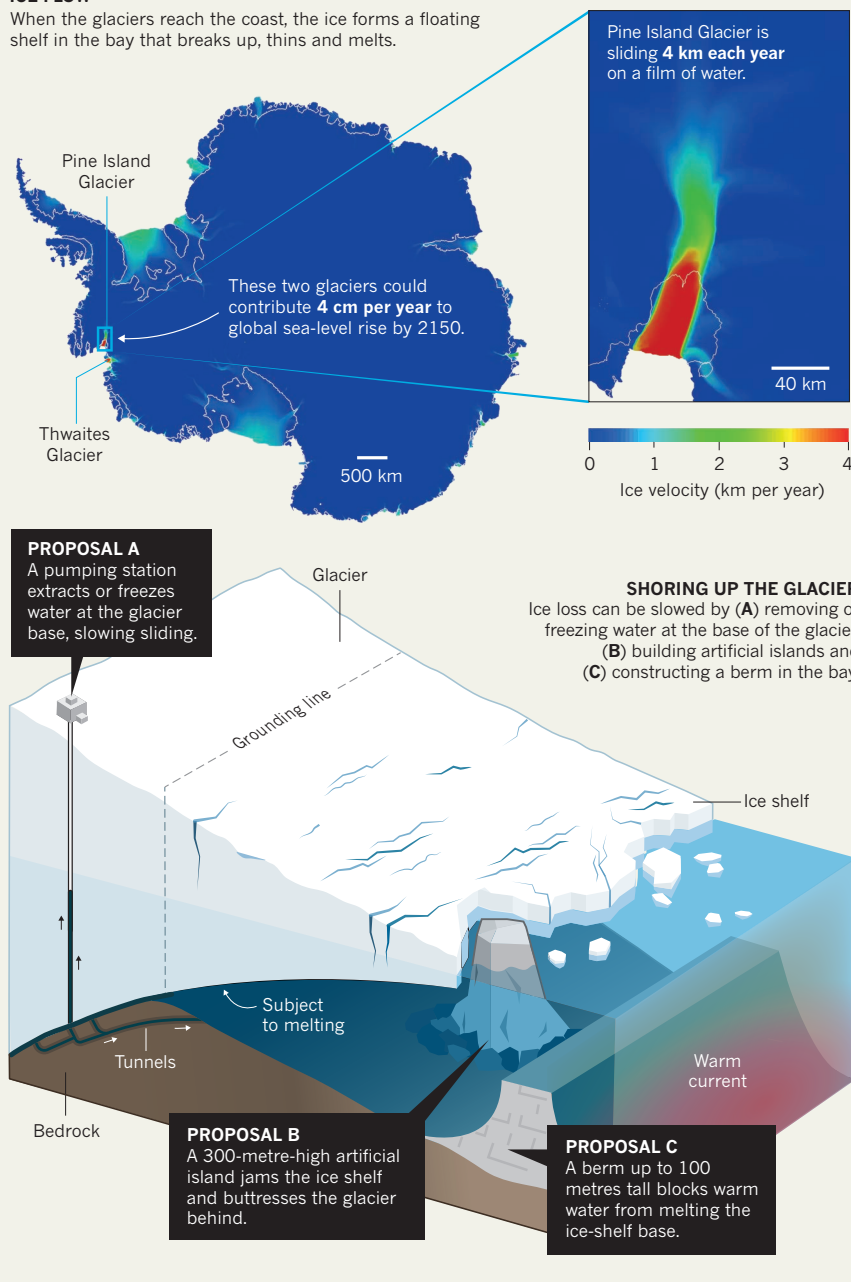
Deeper subglacial water in Antarctica is under pressure and should drain to the ocean without pumping. It could also be frozen by circulating cooled brines beneath the 10-metre-thick layer of sediment scoured at the glacier's base. The Pine Island Glacier might be reached through the nearby volcanic

GLACIAL GEOENGINEERING

Two fast-moving glaciers in West Antarctica — Pine Island and Thwaites — are shedding most of the ice lost from the continent into the sea. Slowing them down could delay global sea-level rise by centuries.

ICE FLOW

When the glaciers reach the coast, the ice forms a floating shelf in the bay that breaks up, thins and melts.



consequences might arise. For instance, if water at a glacier's base is trapped in pockets, some parts of the glacier or ice stream might speed up rather than slow down.

Implementation would require global consent. Antarctica is governed by the Antarctic Treaty, so research there is undertaken within the multilateral framework of the Scientific Committee on Antarctic Research, which meets this June. Countries finance research on the basis of their interests, and a few could take a lead. For example, researchers in China are preparing a \$3-billion plan for polar research in the next decade that includes addressing the feasibility of targeted geo-engineering schemes such as ours. Options for building a research base in the Hudson Mountains, to access the glaciers flowing into the Amundsen Sea, should be discussed.

Around Greenland, sea levels will fall as ice is lost from its interior, reducing the gravitational pull of the ice sheets. This could be as inconvenient for coastal communities as rising seas. There might be mutual benefits to collaboration between Greenlanders and those who are most at risk of rising sea levels, for example in the small island states of Tuvalu or the Maldives.

Geoengineering of glaciers will not mitigate global warming from greenhouse gases. The fate of the ice sheets will depend on how quickly we can reduce emissions. If emissions peak soon, it should be possible to preserve the ice sheets until they are again viable. If they keep rising, the aim will be to manage the collapse of the ice sheets to smooth the rate of sea-level rise and ease adaptation. ■

John C. Moore is chief scientist at the College of Global Change and Earth System Science, Beijing Normal University, China; and professor of climate change at the Arctic Centre, University of Lapland, Rovaniemi, Finland. **Rupert Gladstone** is a geoscientist at the Arctic Centre, University of Lapland. **Thomas Zwinger** is an application scientist at the CSC-IT Center for Science, Espoo, Finland. **Michael Wolovick** is a glaciologist at Princeton University, Princeton, New Jersey, USA.
e-mail: john.moore.bnu@gmail.com

1. Jevrejeva, S., Jackson, L. P., Riva, R. E. M., Grinsted, A. & Moore, J. C. *Proc. Natl Acad. Sci. USA* **113**, 13342–13347 (2016).
2. Hinkel, J. et al. *Proc. Natl Acad. Sci. USA* **111**, 3292–3297 (2014).
3. Favier, L. et al. *Nature Clim. Change* **4**, 117–121 (2014).
4. Joughin, I., Smith, B. E. & Medley, B. *Science* **344**, 735–738 (2014).
5. Joughin, I. et al. *J. Glaciol.* **55**, 245–257 (2009).
6. Joughin, I., Abdalati, W. & Fahnestock, M. *Nature* **432**, 608–610 (2004).
7. Cook, A. J. & Vaughan, D. G. *The Cryosphere* **4**, 77–98 (2010).
8. DeConto, R. & Pollard, D. *Nature* **531**, 591–597 (2016).
9. Mercer, J. H. *Nature* **271**, 321–325 (1978).
10. Joughin, I., Smith, B. E. & Holland, D. M. *Geophys. Res. Lett.* **37**, L20502 (2010).

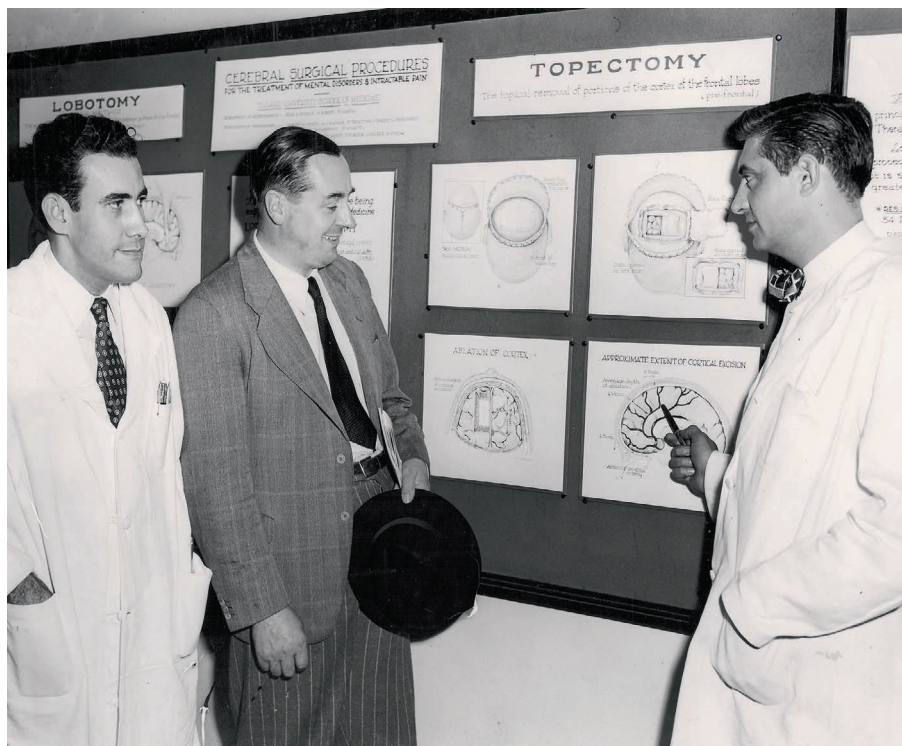
outcrops of the Hudson Mountains. These lie within 80 kilometres of the glacier and the coast, and would be a good base for research into the sub-glacial environment and ice shelves. Again, the costs of such projects appear comparable to those of other large energy and civil-engineering works.

FEASIBILITY TRIALS

Glaciologists and engineers should establish the scientific viability of these projects through fieldwork and computer modelling. The glaciers concerned need extensive

study, including mapping the geomorphology of their beds and the rates at which they are melting. More observations are needed of the North Atlantic's flow onto the Greenland shelf. Climate models need to do a better job of simulating the Southern Ocean.

Potential risks, especially to local ecosystems, need careful analysis. In our view, however, the greatest risk is doing nothing — or if the interventions don't work. The impacts of construction would be dwarfed locally by the effects of the ice sheet's collapse, and globally by rapid sea-level rise. Unexpected



Robert Heath, far right, conducted experiments with brain stimulation.

NEUROSCIENCE

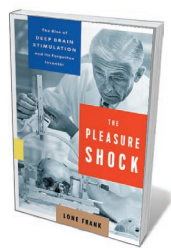
Pulse of the past

Christian Lüscher considers an alarming account of the early days of deep-brain stimulation.

Many people consider deep-brain stimulation (DBS) to have begun in 1987 in Grenoble, France, when Pierre Pollak and Alim Benabid stopped a person's tremor by delivering high-frequency pulses of electricity to her thalamus. In fact, more than three decades earlier, a psychiatrist called Robert G. Heath at Tulane University in New Orleans, Louisiana, had experimented with this approach. Now, science writer Lone Frank pulls Heath (1915–99) from obscurity for her exploration of DBS, *The Pleasure Shock*.

Frank has traced and interviewed surviving patients, former collaborators, family members and current DBS scientists. The result is a rarity: a thrilling, well-researched read. Above all, it is a chilling reminder of how early neurosurgical experimentation knew few ethical boundaries — even firmly within the medical and academic establishment. Heath was chair of Tulane's psychiatry and neurology department for 31 years, from 1949 to 1980.

Today, DBS is an approved treatment for Parkinson's disease, dystonia (uncontrollable



The Pleasure Shock: The Rise of Deep Brain Stimulation and Its Forgotten Inventor
LONE FRANK
Dutton: 2018.

muscle contractions) and essential tremor. Other indications, such as therapy for obsessive-compulsive disorder, depression and addiction, are the focus of intensive research. Just a few patients are treated 'off label', with mixed results.

Heath explored the effects of electrical stimulation on various people, with and without their informed consent. Infamously, in the 1970s he subjected a homosexual man to DBS — for so-called 'conversion therapy', the now-discredited practice of attempting to alter sexual preference. Heath delivered pulses to the septal region of the man's brain (normally active during pleasure), and showed him pornographic material featuring women. Heath obtained permission from a state court to pay a woman to perform sexual acts with the man,

while recording activity in his septum. Heath claimed that the man became heterosexual; this was later contested.

As early as the 1950s, Heath also implanted electrodes in people with schizophrenia, violent behaviour and depression. Towards the end of his career, he proposed stimulation of the cerebellum for the treatment of epilepsy, which he thought was closely related to schizophrenia.

Frank paints a detailed picture of the staggering ethical vacuum in which this egregious research was conducted. This was the period after the rise and fall of pre-frontal lobotomies, from the 1940s to the 1950s. It was also when chlorpromazine was introduced as the first drug for psychosis. Several of the people Heath experimented on experienced post-operative complications, and contemporaries raised serious questions in the literature of the time about his methods and conclusions. All of this makes Heath's long and troubling career a useful subject of study, because it illustrates the beginnings of biological psychiatry.

A strength of the book stems from the parallels that Frank draws with current brain research, notwithstanding today's dramatically different protections for the dignity, rights and welfare of research participants. For example, Frank compares Heath's involvement in the CIA's mind-control experiment MKUltra, which ran from the 1950s to the 1970s, with the current DBS initiative led by the Defence Advanced Research Project Agency, which aims to understand the workings of the brain. The comparison might not please everyone, but it underlines the need for a debate on the independence of research in psychiatry.

Beyond DBS, Heath tested pharmacological substances such as bulbocapnine, which induces therapeutic stupor, on inmates of the Louisiana State Penitentiary at Angola, in the context of MKUltra. Sadly, he was not alone in using prisoners for medical research; from the 1940s to the 1960s, covert testing of drugs, radioactive substances and more was carried out on many people powerless to refuse.

Heath also propounded the idea that a component that he called taraxein, isolated from the blood serum of people with schizophrenia, would elicit symptoms in healthy people. The prisoners he injected quickly learned to perform what he expected them to, for example acting out hallucinations. The episode encapsulates both the alarming ethical mores of the day and why observations need to be made under blinded conditions.

There are a few gaps. Frank talked to many DBS experts, but not to Pollak or Benabid, who pioneered the most important use of the technique today, in the subthalamic nucleus. I also wanted to know whether Heath knew of the Nobel-prizewinning work of physiologist Walter Rudolf Hess on

cats' emotional reaction to electrical brain stimulation, carried out in the 1930s. That evidence could have been a stimulus for Heath's experiments in humans. Some anatomical schematics could have helped the non-expert; and I wished for photographs of Heath at work.

Was Heath an out-and-out monster or a deeply flawed visionary? Frank does not shy away from that question. She vividly describes his charismatic, take-charge personality, analysing his work in the context of his time. He called his house outside New Orleans Hedonia (meaning pleasure); hosted lavish parties; and was a gifted tennis player, all of which probably contributed to his social success in the American deep south of the mid-twentieth century. Frank also makes it clear that much of Heath's research — and the academic environment that allowed it — was appalling.

Today, little remains of Heath's science, in part because he did not systematically investigate underlying mechanisms. In the absence of a demonstration of what causes a condition, it is difficult to propose a stimulation protocol that works. When DBS is used for Parkinson's disease, we know that high-frequency stimulation in the subthalamic nucleus alleviates symptoms, even though the underlying cellular mechanism is debated.

"Contemporaries raised serious questions in the literature of the time about his methods and conclusions."

The future of DBS seems bright thanks to optogenetics, the use of light to control the activity of cells. Over the past decade, scientists have teased apart neural dysfunction in animal models of behavioural conditions such as obsessive-compulsive disorder or addiction. New DBS protocols are currently tested in such models. Treatments likely to reach clinics over the few next years are inspired by optogenetic manipulations of cellular mechanisms to restore normal function in specific brain regions.

The main message of Frank's fascinating, horrifying tale is that progress can be made only through research that is scrupulously ethical. Luckily for the patients of today and tomorrow, DBS got a second chance when it was reinvented in 1987. ■

Christian Lüscher is a neuroscientist and neurologist at the University of Geneva, Switzerland. He aims to establish a circuit model for drug addiction that might enable innovative therapeutic approaches, including DBS. e-mail: christian.luscher@unige.ch

GENETICS

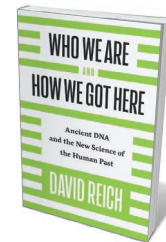
Traces of ancient DNA

Turi King hails David Reich's thrilling account of mapping humans through time and place.

As a field, ancient DNA is paradoxically young — just over 30 years old. And it is booming, thanks to ever-faster sequencing techniques and extraction protocols that can bait specific sections of human DNA out of the vast soup of non-human genetic material in ancient samples. Simultaneously, the field has grabbed the public imagination with findings about the distant past. One such finding was the revelation that people from the Beaker Culture significantly altered Britain's population just 4,500 years ago. Another was the oldest ancient genome ever obtained: that of a 700,000-year-old horse, found in Canadian permafrost, that suggested the ancestor of all today's horses, donkeys and zebras lived some 4 million years ago. I was thrown headlong into the intricacies and difficulties of the field by leading the DNA analysis of the remains of England's King Richard III, discovered under a car park in Leicester in 2012.

Few labs do ancient-DNA work. David Reich's, set up in 2013 at Harvard Medical School in Boston, Massachusetts, was the first in the United States and is one of the most prestigious in the world. It is a juggernaut able to process hundreds of samples a year. Now, with *Who We Are and How We Got Here*, Reich gives us a window into what ancient DNA can tell us about human evolution, the peopling of the world, continent by continent, and the population mixing that makes us who we are today, genetically at least.

Reich's team has developed some of the



Who We Are and How We Got Here: Ancient DNA and the New Science of the Human Past
DAVID REICH
Pantheon: 2018.

most sophisticated statistical and bio-informatics techniques available. Using computers, they painstakingly reconstruct genomic information from fragments of DNA from ancient individuals. They then drill down in search of a new understanding of human history.

It was Reich's lab that did the Beaker work of the headlines. Indeed, the group has been involved in many of the big findings in the field over the past decade, and it's these that Reich discusses. For example, their work contributed to the startling discovery that Neanderthals interbred with the ancestors of all modern humans descended from Europeans, Asians and other non-Africans.

His group's involvement in the genetic analysis of the hominins called Denisovans overturned previous findings based on mitochondrial DNA alone. The work showed that Denisovans and Neanderthals were more closely related to each other than to modern humans. The ancestral groups leading to modern humans separated from the population leading to both Denisovans and Neanderthals 770,000–550,000 years ago, pre-dating by some 100,000–400,000 years the split that led to Neanderthals and Denisovans. And it turned out that ancient Denisovan populations and the ancestors of modern New Guineans had interbred as recently as 54,000–44,000 years ago.

Reich also discusses ghosts in our past. Not all of the genetic make-up of ancient and modern humans can be explained by the current archaeological or historical record. Genetic analysis of ancient and modern populations predicts as-yet-undiscovered groups that must have contributed their DNA to future generations. For example, Reich's lab found that Europeans were more closely related to Native Americans than to East Asians, and this couldn't be explained by recent interbreeding. The researchers suggested that another, now-extinct, group of people must have existed more than 15,000 years ago, and contributed DNA both to the populations that led to



A replica Neanderthal skull.

► modern Europeans and to those that led to modern Native Americans. The team named these people Ancient North Eurasians.

No physical proof of this ghost population existed. Then, another group, led by Eske Willerslev, published genome-wide data from a recent find. They fit. The remains of a boy from Mal'ta in Siberia, dated to about 24,000 years ago, became the type specimen for the Ancient North Eurasians: a ghost made, if not flesh, then at least bone (M. Raghavan *et al.* *Nature* **505**, 87–91; 2014). Other ghost populations have been predicted. As each new type specimen is discovered, more pieces of the puzzle slot into place, and researchers can reach even further back in time.

Reich details many other studies: of the phenomenal spread of the Yamnaya from central Europe to Asia's Altai Mountains some 5,000 years ago; of the Andaman Islanders and the populations of India; of ancient remains in North America, such as the 8,500-year-old Kennewick Man.

What his and other labs are uncovering is the tremendous degree to which populations globally are blended, repeatedly, over generations. Gone is the family tree spreading from Africa over the world, with each branch and twig representing a new population that never touches others. What has been revealed is something much more complex and exciting: populations that split and re-form, change under selective pressures, move, exchange ideas, overthrow one another. Genomics and statistics have drawn back the curtain on the sort of sex and power struggles you'd expect in *Game of Thrones*.

Reich also reflects on how his work can be misinterpreted by the public and those outside the field, in a heartfelt section that I can sympathize with. As soon as some genetic discoveries are published, they can become freighted with prejudices and polarized interpretations. We all belong to one species and we are all related. Yet when genetic differences between populations, for instance, are revealed, the media and interest groups can oversimplify and distort. Some pick and choose results to justify personal, and sadly often political or racist, beliefs. Others sweep the differences under the carpet. Yet, as Reich argues, we do need a non-loaded way to talk about genetic diversity and similarities in populations. This book goes some way to starting that conversation. ■

Turi King is professor of public engagement and reader in genetics and archaeology at the University of Leicester, UK. She is leading the whole-genome analysis of the remains of King Richard III. e-mail: tek2@leicester.ac.uk

People collect recyclable material at a dump in Guwhati, India.



INNOVATION

Waste mountain

Subhra Priyadarshini examines the wide-ranging impacts of India's throw-away culture.

In *Waste of a Nation*, an in-depth investigation of India's feeble fight against mountains of consumerist waste, are robust statistics, compelling history and telling case studies. The authors, anthropologist Assa Doron and historian Robin Jeffrey, also throw the occasional philosophical curve ball, such as: "waste is in the eye of the beholder." The result is both beguiling and disturbing.

As Doron and Jeffrey show, waste in India has generated a vast recycling culture — a world apart, of *kabaadiwalas* (garbage buyers), scavengers and 'rubbish rajas'. The authors reveal the complex cultural, social, political and religious hurdles that

hamper the country's struggle with waste, from unjust pressure on 'low-caste' Dalits to unenforced environmental regulations.

Meanwhile, the mountain builds by an average 100,000 tonnes a day — a fraction of the US tally, but problematic nevertheless. India has few mechanisms for dealing with sewage and hazardous, wet, medical or electronic waste. And, like many other countries, it is losing the battle with megamounds of plastic. Until 1985, the country did not even have an urban-development ministry.

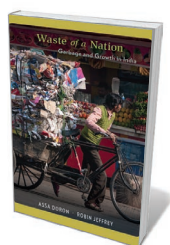
Municipal bodies are responsible for managing waste. But tradition — and the



ANUWAR HAZARIKA/REUTERS

labour-intensive nature of the sector — means that unorganized waste-pickers do most of the dirty work. Their job is to collect and segregate household, commercial and industrial waste for processing in centres where it is sorted for composting, recycling or energy generation. However, the reality rarely reflects this orderly progression.

Non-compliance is rife: waste is often not sorted at source. Ultimately, around 90% of unsorted waste is thrown into dumps. Meanwhile, the millions of scavengers and cleaners are not part of any organized waste-management system, and lack health, safety and legal cover. They face harrowing occupational hazards. Urban dumps in megacities such as Mumbai, Delhi and Kolkata can be clogged with excrement, rotten food, and liquid and solid household wastes that can promote infectious diseases and attract flies, rats and other vectors. Dumps can catch fire; burning tyres, for instance, emit volatile organic compounds and particulate matter. Doron and Jeffrey cite a suspected outbreak of bubonic plague in Surat in 1994 as an example of the breakdown of civic waste management. They point, too, to a community



**Waste of a Nation:
Garbage and
Growth in India**

ASSA DORON & ROBIN
JEFFREY
Harvard University
Press: 2018.

in the brass-working centre Moradabad who extract metals from electronic waste. The illegal operation suffuses their lungs with metallic dust and chemical fumes, and chokes the nearby rivers with mercury and arsenic. In these lands of waste, humans end up being treated as waste.

Sewage, as *Waste of a Nation* underlines, is a prime concern

in a country where more than 560 million people defecate in the open. In 2014, the government of Prime Minister Narendra Modi set out to tackle the problem with the Swachh Bharat ('clean India') campaign, pledging to build 120 million toilets across rural India by 2 October 2019 — the 150th anniversary of Mahatma Gandhi's birth. In 2017, the project achieved a remarkable 70% coverage of rural areas. The sanitation

gap also inspired the 2017 Hindi film *Toilet: Ek Prem Katha*, directed by Shree Narayan Singh. Significant challenges remain, however. Untreated sewage is choking the mighty Yamuna river and parts of the lake system around Bangalore, for instance.

Despite India's tradition of frugality, the rise of consumerism contributes to these issues. The dark side of the economic liberalization of 1991 is the generation of new waste from mines, factories and industrial agriculture. The gradual switch from natural, biodegradable materials to plastics is changing behaviour even among the rural poor. For instance, twigs (*daatuun*) of the medicinal neem tree (*Azadirachta indica*), once used to brush teeth, have given way to plastic tooth-brushes. The latter are a recycling nightmare: separating bristles from the handle is labour-intensive and unrewarding.

Doron and Jeffrey also discuss India's waste market. The world's largest 'ship-breaking' industry is in Alang. Here, retired ships are imported and dismantled, and their parts and materials — primarily steel — are sold for profit. India is also a leading exporter of hair, a market worth almost US\$400 million. Many Hindus have their hair cut in temples to demonstrate devotion, and much of the waste hair is sent to China to be made into wigs.

Doron and Jeffrey analyse the isolated, small-scale attempts of large Indian companies such as ITC and the Ramky Group to recycle waste profitably as well as hygienically, through state-of-the-art containment, neutralization and disposal technologies. For instance, Ramky's first project in 2000 was managing medical wastes for disposal at government-approved centres. By 2016, the country had just 198 approved disposal centres for more than 169,000 hospitals and clinics.

The authors rightly call for a sustainable system. To be practical, this must be motivated by profit, discipline, need, pride or better still, a combination of these. In 2013, China signed up to a 'circular economy' model devoted to recycling as much as possible. This is one approach to sustainability. But India has, as *Waste of a Nation* emphasizes, other strengths that could unite municipalities and individuals. One is its 40,000 civic organizations and action groups that could catalyse coalitions between *kabaadiwalas*, professionals, scientists, engineers, ethical businesses and, importantly, politicians. ■

Subhra Priyadarshini is chief editor of Nature India.



The film imagines Ötzi setting out on a fatal mission to avenge the slaughter of his relatives.

PALAEONTOLOGY

A big-screen requiem for Ötzi the Iceman

Josie Glausiusz is gripped by a portrayal of the Neolithic man whose mummified body was found in the Alps.

Huddled in a gloomy cave, a small group of people wearing hides, their hair bedraggled, mourn a woman who has died in childbirth. “Pitamei, Pitamos,” intones Kelab, a bearded, weather-beaten elder, sprinkling herbs over the body. “Bala,” the group responds, as Kelab’s partner Kisis clasps the dead woman’s bundled baby.

That funeral is one of many moments of tenderness in *Der Mann aus dem Eis* (Iceman). This extraordinary film imagines the life of Ötzi (renamed Kelab), the man whose frozen, mummified, 5,300-year-old corpse was found in 1991 in a mountain pass in the Ötztal Alps, a range that runs along the border between Austria and Italy. All was not altruistic in this world, however, as Berlin-based film-maker Felix Randau emphasizes. In 2001, researchers revealed that Ötzi had been killed when the subclavian artery in his left shoulder was severed by an arrow; and the film features many scenes of horrific violence. The visceral mix of contrasting behaviours humanizes the

protagonists. Randau’s feat is to put flesh and blood and feelings into this ancient, frozen human, recreating a society and even a language that are long gone.

Iceman takes us into the stunning beauty of the South Tyrol a few kilometres from where Ötzi was found. His refrigerated mummy is now preserved in the South Tyrol Museum of Archaeology in Bolzano, Italy. For more than two decades, scientists have analysed Ötzi’s bones, teeth, blood, DNA, stomach contents, intestinal parasites, clothing, shoes, tools and tattoos.

They know his height and weight, that he lived during the Copper Age in the late Neolithic era, owned a valuable copper axe, and was about 45 when he was murdered. In addition to his fatal

Der Mann aus dem Eis (Iceman)

DIRECTOR: FELIX

RANDAU

Port Au Prince Film

& Kultur Produktion:

2018.

arrow injury, and a severe head injury that might indicate a fall, he had a deep, unhealed cut on one hand; this suggests that shortly before his

death, he fended off an attacker who was wielding an axe or knife.

The film sticks closely to archaeological details. Kelab (Jürgen Vogel) lives in a hamlet with Kisis (Susanne Wüst) and other members of his clan, who tend domesticated pigs and goats. He wears a bearskin cap

“Randau’s feat is to put flesh and blood and feelings into this ancient, frozen human, recreating a society and even a language that are long gone.”

and goat-leather leggings. In the tranquillity of early summer, three marauders suddenly appear.

They rape and kill Kisis, slaughter the rest of the inhabitants and burn the village, stealing, as they leave, the group’s

sacred totem (a mirror in a wooden box). Kelab returns from hunting to find carnage. Grief-stricken, he conducts funeral rites for Kisis and goes away, leading a goat, bent on revenge and carrying his orphaned baby in a makeshift sling.

The film’s minimal dialogue is conducted in reconstructed Rhaetic, a language spoken in the eastern Alps in pre-Roman and Roman times. But it’s the action — and Vogel’s wonderfully expressive face — that moves the story along. In pursuit of the marauders, he climbs snow-covered mountainsides in hay-stuffed shoes, and feeds the baby direct from the goat’s teat. He later leaves the child with a young woman and her aged father.

In its portrayal of women, however, the movie presents only a narrow picture of the known archaeological record. Women are mostly shown engaging in child-rearing and sex. Recent research suggests that Neolithic women were athletic and versatile, grinding grain and processing milk, meat, hides and wool.

One study, led by Alison Macintosh at the University of Cambridge (A. Macintosh *et al. Sci. Adv.* 3, eaao3893; 2017), looked at the skeletal remains of women in Central Europe from about 5,300BC to AD850. The researchers found that the arm bones of the Neolithic women were 11–16% stronger for their size than those of top female rowers today. But the film’s depiction of rape as a weapon of war remains tragically resonant.

Writing about Ötzi for *Discover* magazine in early 2002 (see go.nature.com/2fdjt0j), I developed an affinity for this messenger from a distant time. As I watched him in the film, dying in the deep snow, I felt a twinge of sadness. Pitamei, Pitamos, I wanted to say. Bala. ■

Josie Glausiusz is a science journalist in Israel. Twitter: @josiegz
e-mail: josiegz@gmail.com

Correspondence

Social justice in the Belt and Road plan

Guo Huadong outlines proposals to share big data on Earth observations across Asia, the Middle East and East Africa to ensure that China's Belt and Road Initiative (BRI) will contribute to sustainability (*Nature* **554**, 25–27; 2018). Sustainability also includes issues of social justice and so, on the basis of the United Nations Sustainable Development Goals, the BRI should include respect for human rights as well, so that everyone benefits.

In our view, the BRI could achieve this social sustainability by including bottom-up participation in all countries involved. China should do more to encourage European countries to sign up to the BRI and enhance the balance of views.

Transnational 'digital democracy' should reach beyond the exchange of goods, services and investments by empowering citizens, increasing opportunities for all and promoting peace (D. Helbing and P. Seele *Nature* **549**, 458; 2017).

Peter Seele *University of Lugano (USI), Lugano, Switzerland.*

Dirk Helbing *ETH Zurich, Switzerland.*
peter.seele@usi.ch

Grounding ME/CFS therapies in science

I welcome the suggestion that patients with myalgic encephalomyelitis (also known as chronic fatigue syndrome; ME/CFS) should not be dismissed (M. Sharpe *et al.* *Nature* **554**, 31; 2018). However, as someone who has been diagnosed with ME/CFS for 25 years, I contend that this argument should not be misused to perpetuate ineffective therapies that could raise false hopes and might amount to mistreatment.

As you point out (*Nature* **553**, 14–17; 2018), the PACE trial authors (including two co-authors of Sharpe *et al.* in *Nature*) and others promote a form of

cognitive behavioural therapy that assumes ME/CFS symptoms can be reversed by teaching people to think differently, and a prescribed form of graded exercise that might be harmful.

Sharpe and colleagues urge readers not to reject scientific evidence that supports the use of such approaches. However, the Cochrane Reviews they cite rely on the results of the disputed PACE trial and several other studies that have similar methodological flaws. It is also notable that Sharpe and colleagues concluded: "There was little evidence of differences in outcomes between the randomised treatment groups at long-term follow-up" (see M. Sharpe *et al.* *Lancet Psychiatry* **2**, 1067–1074; 2015).

The returns might be some way off, but the latest moves to pursue the growing evidence that ME/CFS symptoms are rooted in pathology is the proper approach (see, for example, *go.nature.com/2fimftx*).

Robert Saunders *Balcombe, West Sussex, UK.*
rhsaunders@gmail.com

Bring supplementary citations into view

As David Shotton argues, all publishers should make bibliometric citations free to access, analyse and reuse (*Nature* **553**, 129; 2018). I was disappointed, therefore, to find that references in online-only supplements in *Nature* can still be invisible, even though the problem was raised ten years ago (F. Seeber *Nature* **451**, 887; 2008).

The use of online-only supplements and the number of citations they host has been rising steeply. For example, a highly cited paper entitled 'Worldwide acceleration of mountain erosion under a cooling climate' (F. Herman *et al.* *Nature* **504**, 423–426; 2013) uses a global data set compiled by mining data from more than 400 publications. These references are listed only

in the paper's Supplementary Information and are invisible to Google Scholar and other citation-metric websites. One of those publications was mine.

As long as citation metrics are used for performance evaluation and to measure impact, lost bibliographic information is damaging — especially for early-career researchers. It is high time that *Nature* implemented measures to ensure the transparency Shotton advocates. **Kalin T. McDannell** *Geological Survey of Canada, Calgary, Canada.*
kalin.mcdannell@canada.ca

Editor's note — *Nature* has now instigated a review of referencing practices, with the intention that all citations should be appropriately visible and indexed.

Nobel nominations reveal a way to win

We analysed hundreds of nominations for the Nobel Prize in Physiology or Medicine for the period 1901–66 to gain insight into how Nobel committees judged scientific research as worthy of this ultimate accolade.

Nomination letters become public 50 years after the award. Some simply highlight their nominees' lifetime achievements and portray individuals as world-leading scholars in their field. The (unsuccessful) nomination letters for antiseptics pioneer Joseph Lister, for instance, stated that he had "done more for the good of humanity than any other [living] member of the medical profession".

Others focus on discoveries that could open up new research areas. The sponsors of physician Charles Huggins, for example, argued that his "visionary" work meant that cancer was no longer perceived as an insurmountable problem. Huggins was awarded the Nobel in 1966 for his work on the hormonal treatment of prostate cancer.

The outcome of the nominations we studied

suggested to us that Nobel committees awarding this prize over the period in question were motivated by the potential research impact of a single innovation, rather than by a distinguished research record. This might explain why the pioneers of anaesthesia did not get the award (N. Hansson *et al.* *Anesthesiology* **125**, 34–38; 2016). **Nils Hansson, Thorsten Halling, Heiner Fangerau** *Medical Faculty, Heinrich Heine University Düsseldorf, Germany.*
nils.hansson@hhu.de

Peer-review panels can be a wolf pack

I agree with Gemma Derrick that grant-review panels should be more collaborative so that they benefit from all the expertise around the table (*Nature* **554**, 7; 2018). What I've seen of how these panels are conducted brings to mind a predatory wolf pack, rather than the tug of war between alliances that she describes.

My experience over the years is that one or two internationally eminent alpha personalities lead the room with their views on the applicant's kudos and calibre and whether this is sufficient to warrant consideration. The group dynamic then shifts palpably as panellists fall into line to concur. Individuals who would normally review work in their field confidently and independently start to act as a mob.

This unproductive behaviour would be alleviated if panels required all members to make their own initial assessment of proposals ahead of the meeting and to submit a voice message summarizing their appraisal to the chair. Final scoring could then be based on a composite of the raw scores by individual panel members, further informed by any extra insights arising from the subsequent group discussion.

Daniel Altmann *Imperial College London, UK.*
d.altmann@lms.mrc.ac.uk

Oxidation softens mantle rocks

Seismic waves that propagate through a layer of Earth's upper mantle are highly attenuated. Contrary to general thinking, this attenuation seems to be strongly affected by oxidation conditions, rather than by water content. [SEE LETTER P.355](#)

TETSUO IRIFUNE & TOMOHIRO OHUCHI

The outermost layer of the solid Earth is divided into tectonic plates that move around on a region of the upper mantle called the asthenosphere. Seismic waves from earthquakes travel through the asthenosphere at relatively low speeds, and are highly attenuated as a result of energy dissipation, a property known as anelastic behaviour¹. These seismic characteristics are usually associated with low viscosity (a 'softening') in the mantle rock peridotite. On page 355, Cline *et al.*² demonstrate for the first time that this softening is influenced by oxidation conditions — a result that could have major implications for our understanding of the upper mantle.

The softening of peridotite in the asthenosphere was initially thought to be caused by small amounts of melted material³. Such material would act as a lubricant between crystals of the mineral olivine that are abundant in peridotite (Fig. 1). However, in the 1990s, it was shown that this effect is limited to particularly warm regions of the mantle, such as beneath the volcanoes that occur along mid-ocean ridges, where melted material is abundant enough to form interconnected networks^{4,5}.

Over the past two decades, the consensus has been that the presence of water leads to substantial softening of peridotite⁶. Experiments that measure the deformation of olivine crystals under large strains^{7–9} have shown that small amounts of water can enhance both the sliding of grain boundaries (the interfaces between crystals) and the deformation of individual crystals. However, the absence of appropriate equipment and techniques has meant that there have been no experiments to assess the anelastic behaviour of peridotite when it has a realistic water content (up to a few hundred parts per million¹⁰) and is under the small strains associated with seismic-wave propagation.

Cline and colleagues used a sophisticated method¹¹ to subject aggregates of olivine to high temperatures and pressures, and to oscillations that mimic seismic waves known as shear waves. The authors considered aggregates that had a range of water contents and oxidation states, similar to those expected



Figure 1 | Mantle peridotite from San Carlos, Arizona. Peridotite is the dominant rock in the upper part of Earth's mantle. It consists mainly of the mineral olivine (light green), with smaller amounts of other minerals such as pyroxene, spinel and garnet (darker colours). Cline *et al.*² subjected aggregates of olivine to high temperatures and pressures, and to oscillations that mimic seismic waves known as shear waves. They discovered that the speed and attenuation of the waves were insensitive to the water content of olivine, but strongly dependent on oxidation conditions — findings that could reshape our view of the upper mantle.

in the asthenosphere. They discovered that the speed and attenuation of the waves were insensitive to water content, in contrast to expectations from the results of the large-strain deformation experiments^{7–9}.

Instead, Cline *et al.* found that the seismic properties of their olivine aggregates were markedly dependent on oxidation state: wave speed decreased and attenuation increased with increasing oxygen fugacity (degree of oxidation). This finding could imply that the low speeds and high attenuation of seismic waves in the asthenosphere, particularly above sinking (subducting) tectonic plates, are partly caused by the highly oxidized conditions that are expected in such regions.

To explain their results, the authors suggest that ferric iron (Fe^{3+}) and associated metal-ion vacancies that exist in olivine become stabilized under oxidized conditions, yielding high concentrations of crystal defects and/or a modified grain-boundary structure. Such changes are expected to enhance the rate at which defects diffuse through the crystals,

leading to the observed anelastic behaviour.

Because olivine is the most common mineral throughout the upper mantle, Cline and colleagues' findings could have implications for the oxygen fugacity and water content of not only the asthenosphere but also the entire upper mantle. For instance, if oxygen fugacity were entirely responsible for the seismic attenuation, it would have to fall by a factor of about 100 between the asthenosphere and the underlying part of the upper mantle to account for the observed decrease in attenuation¹. Such a drop is consistent with petrological data from deeper regions of the upper mantle^{12,13}.

By contrast, if attenuation were attributable to water alone, water content would need to decrease by a factor of about 100 between the asthenosphere and the underlying layer, on the basis of an earlier model⁶ and observations¹, and peridotite at the base of the upper mantle would be almost completely dry. This prediction conflicts with observations of the mantle's electrical conductivity, which is sensitive to water content. Electrical-conductivity profiles

are either roughly constant or increase with depth throughout the upper mantle below the asthenosphere^{14–16}, suggesting that water content should follow similar trends. This enigma can be solved if water is not the primary cause of seismic attenuation, as shown by Cline and colleagues.

Nevertheless, there are some issues regarding the applicability of Cline and colleagues' results to the actual mantle. For instance, the authors artificially increased the water content of some of their olivine aggregates using a technique called doping, in which a trace amount of one element is substituted for another. This process introduced artificial crystal defects whose mobility might differ from the defects intrinsic to olivine — although the authors argue that these artificial defects do not affect their conclusions. The effect of oxygen fugacity on the mobility of these different types of defect is also unknown.

Future studies on the seismic properties of olivine could avoid the need for doping by subjecting aggregates to higher pressures than those used by Cline and colleagues. For example, measurements could be made using an oscillation technique that combines a large-volume press and X-ray observations¹⁷. Future

experiments should include wider ranges of oxidation conditions and olivine grain sizes than those considered by Cline *et al.*, to confirm the dominance of oxygen fugacity over other causes of anelastic behaviour.

Although some petrological evidence suggests that oxygen fugacity in the mantle generally decreases with depth^{12,13}, it has been difficult to evaluate how such oxidation states vary laterally. The probable link between oxygen fugacity and attenuation of seismic waves in peridotite could enable 3D mapping of oxidation states in the deep mantle, using data obtained with an imaging technique called seismic tomography. Meanwhile, the lack of correlation between water content in mantle olivine and seismic attenuation, if confirmed by independent studies at higher pressures, might require scientists to reconsider the role of water in the softening of mantle rocks, and the distribution and circulation of water throughout the deep Earth. ■

Tetsuo Irifune and Tomohiro Ohuchi are in the Geodynamics Research Center, Ehime University, Matsuyama 790-8577, Japan. T.I. is also at the Earth Life Science Institute, Tokyo Institute of Technology.

e-mails: irifune@dpc.ehime-u.ac.jp; ohuchi@sci.ehime-u.ac.jp

1. Dalton, C. A., Ekström, G. & Dziewoński, A. M. *J. Geophys. Res. Solid Earth* **113**, B09303 (2008).
2. Cline, C. J. II, Faul, U. H., David, E. C., Berry, A. J. & Jackson, I. *Nature* **555**, 355–358 (2018).
3. Anderson, D. L. & Spetzler, H. *Phys. Earth Planet. Inter.* **4**, 62–64 (1970).
4. Wilcock, W. S. D., Solomon, S. C., Purdy, G. M. & Toomey, D. R. *Science* **258**, 1470–1474 (1992).
5. Hirth, G. & Kohlstedt, D. L. *J. Geophys. Res. Solid Earth* **100**, 15441–15449 (1995).
6. Karato, S. & Jung, H. *Earth Planet. Sci. Lett.* **157**, 193–207 (1998).
7. McDonnell, R. D., Peach, C. J. & Spiers, C. J. *J. Geophys. Res. Solid Earth* **104**, 17823–17845 (1999).
8. Ohuchi, T. *et al. Sci. Adv.* **1**, e1500360 (2015).
9. Karato, S., Paterson, M. S. & FitzGerald, J. D. *J. Geophys. Res. Solid Earth* **91**, 8151–8176 (1986).
10. Hirschmann, M. M. *Annu. Rev. Earth Planet. Sci.* **34**, 629–653 (2006).
11. Jackson, I. & Paterson, M. S. *Pure Appl. Geophys.* **141**, 445–466 (1993).
12. McCammon, C. & Kopylova, M. G. *Contrib. Mineral. Petrol.* **148**, 55–68 (2004).
13. Woodland, A. B. & Koch, M. *Earth Planet. Sci. Lett.* **214**, 295–310 (2003).
14. Sarafian, E. *et al. Geochem. Geophys. Geosyst.* **16**, 1115–1132 (2015).
15. Utada, H., Koyama, T., Obayashi, M. & Fukao, Y. *Earth Planet. Sci. Lett.* **281**, 249–257 (2009).
16. Yoshino, T., Matsuzaki, T., Yamashita, S. & Katsura, T. *Nature* **443**, 973–976 (2006).
17. Li, L. & Weidner, D. J. *Rev. Sci. Instrum.* **78**, 053902 (2007).

tissue⁵, and a third used radiocarbon dating to identify hippocampal-neuron turnover⁶. However, methodological challenges make human studies difficult to interpret, and more are required to make definitive conclusions.

Sorrells *et al.* set out to address this need using classic immunohistochemical techniques in which specific antibodies are bound to proteins of interest, revealing their locations in tissue. The authors used this strategy to count neural precursor cells, proliferating cells and immature neurons in samples from 59 human subjects, spanning fetal development through to old age (Fig. 1). They found streams of all three cell types migrating from an embryonic 'germinal zone' to the developing dentate gyrus at 14 weeks of gestation. By 22 weeks, migration was reduced, and immature neurons were largely restricted to the dentate gyrus. And there were many fewer immature neurons at one year of life than at earlier stages. The oldest sample containing immature neurons was taken from a 13-year-old individual. These findings are in stark contrast to the prevailing view that human hippocampal neurogenesis extends throughout adult life.

Is it possible to reconcile the findings with previous human data? Although direct comparisons are difficult, Sorrells *et al.* offer some explanations. For example, they find that DCX and PSA-NCAM, two proteins that reliably mark immature neurons in animals, can label mature neurons and non-neuronal glial cells in humans. Indeed, the authors show that these two markers unambiguously identify immature neurons only if both are expressed in a

NEUROSCIENCE

Questioning human neurogenesis

Neurons are born in the brain's hippocampus throughout adulthood in mammals, contributing to the region's functions in memory and mood. But a study now questions whether this phenomenon really extends to humans. SEE LETTER P.377

JASON S. SNYDER

Pick up any article on neuronal development in adulthood, and there is a good chance you will read that the birth of new neurons has been observed in the hippocampal region of the brain in every mammalian species examined, including humans. This idea underlies the view — widespread among neuroscientists — that analysis of such neurogenesis in animals can benefit our understanding of learning, emotional disorders and neurodegenerative disease in humans. But Sorrells *et al.*¹ report on page 377 that, unlike in other mammals, the last new neurons in the human hippocampus are generated in childhood. These findings are certain to stir up controversy.

Today, it is common knowledge that the brain can change according to needs and demands. But it was not always so. In the 1960s, biologist Joseph Altman reported

that new neurons are generated in the adult brain, specifically in a hippocampal subregion called the dentate gyrus, which is now known to be crucial for memory². Further research languished, however, owing to scepticism about the brain's capacity for such dramatic plasticity. It wasn't until the 1990s, with the development of improved techniques for visualizing brain cells, that acceptance of adult neurogenesis became widespread³.

Although the scope and function of neurogenesis remain debatable, there has been a general consensus that the hippocampus is one region in which adult neurogenesis exists in humans as it does in animals. This is based on several studies. For example, one study in patients given a synthetic nucleoside molecule called bromodeoxyuridine (BrdU) showed that it had been incorporated into the DNA of dividing cells in the dentate gyrus⁴. Another found that protein markers of neurogenesis in animals were present in post-mortem human brain

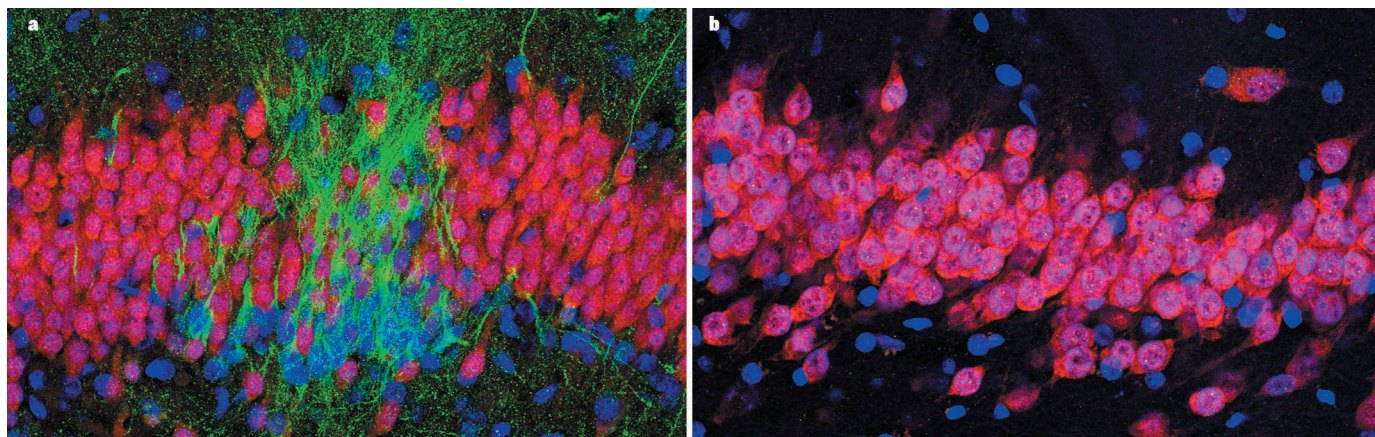


Figure 1 | Decreasing neurogenesis with age. Sorrells *et al.*¹ examined slices of the hippocampus from human brains at various stages of life, to investigate when new neurons are generated. Green indicates the location of the protein DCX, which is produced in new neurons; red indicates the protein NeuN,

which is produced in mature neurons; blue indicates a fluorescent marker called DAPI, which stains all cell nuclei. **a**, At birth, many new neurons can be seen. **b**, By contrast, the authors observed no new neurons in the adult hippocampus.

S. SORELLS ET AL./NATURE

cell. Similarly, the group demonstrated that it is possible to obtain BrdU-like immunohistochemical labelling in tissue that did not actually contain BrdU. Nonspecific labelling could therefore have led to false-positive results in previous studies.

The researchers' careful approach also speaks to the challenges of performing neurogenesis work in humans. Animal studies have shown that PSA-NCAM is modified by previous experiences⁷ and that DCX degrades if tissue is not rapidly preserved⁸. An apparent loss of neurogenesis could therefore reflect changes in marker expression, especially if stringent criteria are used to define new neurons. Given that there are debates about hippocampal precursor-cell identity even in rodents⁹, it is also possible that we simply do not know what to look for in humans.

Sorrells *et al.* minimized these issues in several ways. First, they observed neurogenesis in the hippocampus of infants and children, which served as a positive control. Second, they used a variety of adult samples to minimize the possibility that problems with tissue health or preservation could confound their results. Third, they used diverse markers of neurogenesis to gain multiple lines of evidence. Nonetheless, further investigation will be needed to see whether Sorrells and colleagues' conclusions will stand the test of time.

How do the authors' findings fit with the animal literature? With a bit of conceptual recalibration, they might fit quite well. Rodents are born with relatively immature nervous systems, so adult rodent neurogenesis could be a decent model of neurogenesis in children or adolescents. Given that depression, schizophrenia and Alzheimer's disease are rooted in early hippocampal defects, even neurons generated in childhood could have a key role in the aetiology of disease in humans. In addition, primate data¹⁰ suggest that new neurons in humans could go through an extended period of maturation (years or even decades)

relative to what occurs in rodents, during which time they might have enhanced plasticity and important functional properties. Thus, whereas the continual addition of new neurons might provide plasticity in adult rodents, the prolonged development of neurons could provide a similar plasticity in adult humans.

At the other end of the developmental spectrum, even in rodents, neurogenesis is very low by middle age². Thus, Sorrells and colleagues' human data again are not wholly inconsistent with the animal literature. If the focus of rodent studies were shifted to identifying the mechanisms by which neurogenesis diminishes over time, and to how neurogenesis can be enhanced to offset pathology caused by age and disease, we just might be able to translate the authors' sobering findings into discoveries that improve human health. ■

Jason S. Snyder is in the Department of Psychology, Djavad Mowafaghian Centre for Brain Health, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada. e-mail: jasonsnyder@psych.ubc.ca

1. Sorrells, S. F. *et al.* *Nature* **555**, 377–381 (2018).
2. Altman, J. & Das, G. D. *J. Comp. Neurol.* **124**, 319–335 (1965).
3. Gross, C. G. *Nature Rev. Neurosci.* **1**, 67–73 (2000).
4. Eriksson, P. S. *et al.* *Nature Med.* **4**, 1313–1317 (1998).
5. Knoth, R. *et al.* *PLoS ONE* **5**, e8809 (2010).
6. Spalding, K. L. *et al.* *Cell* **153**, 1219–1227 (2013).
7. Lopez-Fernandez, M. A. *et al.* *J. Neurosci.* **27**, 4552–4561 (2007).
8. Boekhoorn, K., Joels, M. & Lucassen, P. J. *Neurobiol. Dis.* **24**, 1–14 (2006).
9. Bonaguidi, M. A. *et al.* *Cold Spring Harb. Perspect. Biol.* **8**, a018838 (2016).
10. Kohler, S. J., Williams, N. I., Stanton, G. B., Cameron, J. L. & Greenough, W. T. *Proc. Natl Acad. Sci. USA* **108**, 10326–10331 (2011).

This article was published online on 7 March 2018.

CANCER GENOMICS

Landscapes of childhood tumours

Two analyses of the genetic alterations that characterize paediatric cancers reveal key differences from adult cancers, and point to ways of optimizing therapeutic approaches to combating cancer in children. SEE ARTICLE P.321 & LETTER P.371

**PRATITI BANDOPADHAYAY
& MATTHEW MEYERSON**

The mapping of the human genome, followed by the explosion in next-generation genome sequencing, has revolutionized our understanding of cancer. These advances have paved the way for precision-medicine approaches to treating adult cancers. Two papers in *Nature* report the first

pan-cancer genomic analyses in children. In the first, Gröbner *et al.*¹ (page 321) analysed sequences of whole exomes (all the protein-coding regions in the genome) or whole genomes for 961 cancers across 24 tumour types, with an emphasis on tumours of the central nervous system. In the second, Ma *et al.*² (page 371) used similar analyses to characterize 1,699 cancers across 6 types of cancer tissue, particularly leukaemias.

The findings provide valuable insights into the mechanisms that shape the genomes of childhood cancers.

Adult cancers frequently involve multiple genetic alterations that together drive cancer progression, including small mutations of one or a few DNA bases, and larger changes called structural variants that span more than 1,000 bases. Such drivers can be shared across cancer types³. One of the most meaningful outcomes of the current studies is their confirmation that the genomic landscape of childhood cancers differs from this picture. Previous studies of individual paediatric cancer types^{4–7} have revealed that they have fewer mutations and structural variants, on average, than do adult cancers^{3,8}, but the current pan-cancer analyses take this further, systematically highlighting several key differences between childhood and adult cancer genomes (see ‘The differing genomic landscapes of childhood and adult cancers’).

First, there are fewer mutations and structural variants in paediatric cancers than in adult cancers. For instance, Gröbner *et al.* report a mutation rate 14 times lower for childhood than for adult cancers. Furthermore, both groups find that the total number of mutations in paediatric-cancer genomes correlates significantly with age — consistent with the idea that cells accumulate mutations with age.

Second, paediatric cancers are frequently defined by a single driver gene. For instance, 57% of the cancers in Gröbner and colleagues’ analysis harboured single driver mutations. These authors also highlight the fact that germline mutations, which are inherited from parents and are present in all cells of the body, are a causative factor in childhood cancers — 7.6% of cancers in the authors’ cohort are associated with detectable germline mutations. Furthermore, paediatric cancers tend to be enriched in either mutations or structural variants, rather than a mixture of the two. Indeed, the group observes enrichment of germline mutations involved in a DNA-repair pathway called mismatch repair in cancers defined by mutations, and germline mutations in a tumour-suppressor gene, *TP53*, in cancers characterized by structural variants. These differences highlight potential mechanisms by which different paediatric cancer genomes are shaped.

Third, different genes are mutated in paediatric compared with adult cancers. Only 30% of significantly mutated genes identified by Gröbner and co-workers (those that have acquired more mutations than would have been expected to occur by chance, and so are likely to be involved in cancer progression), and only 45% of those reported by Ma and colleagues, overlap with adult pan-cancer analyses. These differences are borne out in the groups’ mutation-signature analyses, which provide information about the mutational processes

THE DIFFERING GENOMIC LANDSCAPES OF CHILDHOOD AND ADULT CANCERS

Two studies^{1,2} have analysed genome sequences from a range of childhood cancers, and uncovered key differences from adult cancers.

Feature	Childhood	Adult
Mutation rate	Lower	Higher
Cancer-driving mutations	Frequently single	Multiple
Mutation specificity	Disease-specific	Shared

that lead to a particular pattern of mutations.

Fourth, and perhaps most intriguingly, driver mutations tend to be specific to individual paediatric cancer types, with minimal overlap across diseases. This is in contrast to adult cancers, which more frequently share mutations across types, according to Gröbner and colleagues’ analysis. This finding by the current studies might reflect the differing paths to cancer development between adult and paediatric cancers. Adult cancers often arise through a multiple-hit process in which alterations in genes generally beneficial to cell survival accumulate as cells become cancerous⁹. By contrast, a model of paediatric cancers posits that, in some cases, a single, specific driver alteration might promote cancer development in certain cell lineages, if it results in

“Different genes are mutated in paediatric compared with adult cancers.”

aberrant gene expression during a crucial period of development⁶. Indeed, a study in mice has also highlighted the importance of developmental context and the timing of genomic perturbation in tumour development¹⁰.

The insights gleaned from the current analyses have implications for precision-medicine approaches for childhood cancers. Gröbner *et al.* found that about 50% of the tumours that they profiled harbour genomic alterations that can be targeted (directly or indirectly) by drugs that are available or under development. This number, which is consistent with previous reports^{11,12}, is a cause for optimism. The findings also provide insights into how clinical assays could be designed to ensure robust detection of alterations specific to paediatric tumours. Assays must profile genes that are significantly mutated across childhood cancers, with sufficient sensitivity to detect single driver alterations in an individual’s genome, and must be specifically designed to include mutations and structural variants in both coding and non-coding regions of the genome.

Furthermore, the studies reinforce the need for paediatric oncologists to consider the high incidence of germline mutations in their patients. Clinicians should offer genetic counselling (in which patients are advised about risks and management options for genetic disorders), testing for germline alterations and appropriate screening for families who are

found to harbour germline mutations.

Although the current studies provide valuable insights, much work is still required to complete the picture. Gröbner and colleagues were unable to identify driver alterations in 10% of tumours, and neither group analysed enough samples in a specific cancer type to detect infrequent mutations. Childhood cancers are, by definition, rare tumours. Continued collaboration and data-sharing are required to amass information from enough tumours of each type to comprehensively identify recurrent driver alterations. Furthermore, given that both groups identified structural variants, which often occur in non-coding regions, whole-genome sequencing is needed to detect drivers outside coding regions. Data from both studies are available for review — Gröbner and colleagues’ at go.nature.com/2bq3oyh, and Ma and colleagues’ at go.nature.com/2svr9hh. This is a key step in paving the way for further analytical efforts across large cohorts of paediatric tumours.

Last but not least, it will be necessary to elucidate the mechanisms by which the identified genetic alterations drive childhood cancer. This will improve our ability to target these alterations therapeutically. ■

Pratiti Bandopadhyay is in the Department of Pediatric Oncology and **Matthew Meyerson** is in the Department of Medical Oncology, Dana–Farber Cancer Institute, Boston, Massachusetts 02215, USA. Both authors are at the Broad Institute of Harvard, Massachusetts Institute of Technology and Harvard Medical School, Boston. e-mails: pratiti_bandopadhyay@dfci.harvard.edu; matthew_meyerson@dfci.harvard.edu

- Gröbner, S. N. *et al.* *Nature* **555**, 321–327 (2018).
- Ma, X. *et al.* *Nature* **555**, 371–376 (2018).
- Lawrence, M. S. *et al.* *Nature* **505**, 495–501 (2014).
- Zhang, J. *et al.* *Nature Genet.* **45**, 602–612 (2013).
- Bandopadhyay, P. *et al.* *Nature Genet.* **48**, 273–282 (2016).
- Lee, R. S. *et al.* *J. Clin. Invest.* **122**, 2983–2988 (2012).
- Crompton, B. D. *et al.* *Cancer Discov.* **4**, 1326–1341 (2014).
- Beroukhi, R. *et al.* *Nature* **463**, 899–905 (2010).
- Greaves, M. & Maley, C. C. *Nature* **481**, 306–313 (2012).
- Han, Z.-Y. *et al.* *Nature Commun.* **7**, 10421 (2016).
- Ramkissoon, S. H. *et al.* *Neuro-Oncology* **19**, 986–996 (2017).
- Worst, B. C. *et al.* *Eur. J. Cancer* **65**, 91–101 (2016).

This article was published online on 28 February 2018.

Waves cornered

The experimental discovery of materials known as higher-order topological insulators corroborates theoretical predictions and expands the toolbox for integrated optics and mechanical devices. [SEE LETTERS P.342 & P.346](#)

MICHEL FRUCHART & VINCENZO VITELLI

When waves encounter an obstacle, they are typically scattered in all directions. But at the edges of materials called topological insulators, waves are topologically protected, which means that they can propagate in spite of structural imperfections. Two papers in *Nature*, by Serra-Garcia *et al.*¹ (page 342) and Peterson *et al.*² (page 346), and a third published on the arXiv preprint server by Imhof *et al.*³, now report experimental evidence for a new type of topological insulator that supports protected waves at its corners, rather than at its edges. Such materials could find applications in the design of waveguides (structures that restrict wave propagation) and in integrated optics and mechanics. More importantly, they are the first confirmation of a theoretical description that could unify observations previously thought to be unrelated in condensed-matter physics⁴.

A good chocolate is hard on the outside, but soft on the inside. Topological insulators

are the opposite. The d -dimensional interior (bulk) of a topological insulator is 'hard' in the sense that it will not react to external stimuli at certain frequencies: there is a range of frequencies, known as a gap, at which waves cannot propagate. By contrast, the $(d-1)$ -dimensional boundaries not only allow wave propagation, but also guarantee the existence of topologically protected oscillations (modes) at the gap frequencies. Such oscillations are localized in dimension $d=1$, for which the boundaries are points, and propagate along the boundaries in $d>1$ (Fig. 1a).

Crucially, the existence of these protected edge modes can be traced to the physics of the bulk material. One can summarize the mathematical description of wave propagation inside a topological insulator as an intricate knot and that outside it as a simple loop. The knot must be cut at the edges to match the 'untwisted' wave propagation outside. Cutting the knot allows modes that have otherwise forbidden frequencies to be present.

In 2017, the theory of topological insulators was extended^{5,6} to include higher-order

examples⁷⁻⁹, such that ordinary topological insulators appear at the first order. A higher-order insulator can be thought of as having a nested topological structure. For example, in second-order topological insulators, the properties of the bulk cause the $(d-1)$ -dimensional boundaries to have frequency gaps. However, the boundaries themselves are topological insulators — protected modes are supported on $(d-2)$ -dimensional corners or hinges (Fig. 1b). In third-order insulators, the boundaries of the boundaries are topological insulators, and protected modes exist on $(d-3)$ -dimensional corners (Fig. 1c). The result of this hierarchical process is more subtle than merely adding together topologically protected edges without the bulk of a higher-order insulator. For instance, in a quadrupole topological insulator (a second-order insulator in dimension $d=2$), there is only one mode in each corner, yet each mode is shared between two edges.

The three current papers report experimental evidence for higher-order topological insulators. More specifically, they identify the topologically protected corner modes associated with quadrupole insulators. The authors achieved this feat using artificial structures called metamaterials, which are engineered to have properties not found in nature¹⁰.

Serra-Garcia *et al.* obtained the required topological structure by tuning the vibrational excitations of connected vibrating plates. Peterson *et al.* used coupled light-trapping devices known as microwave resonators. Finally, Imhof *et al.* used a network of electrical components (capacitors and inductors) that were linked to one other. All three teams showed that the corner modes of their topological insulators exist at frequencies not permitted in the bulk — a clear indication that such modes originate from the bulk's topology. Peterson and colleagues went a step further by explicitly demonstrating the robustness of the corner modes to deformation of the edges.

The theoretical prediction of higher-order systems rests on a generalization of electric dipole moments to multipole moments that are quantized (having only specific discrete values)^{5,6}. Whereas conventional topological insulators are related to dipoles, higher-order insulators are related to quadrupoles, octupoles, and so on. This theory has been corroborated by the authors' experimental realizations of quadrupole systems. However, the experiments did not directly measure the responses of the topological insulators to electromagnetic fields, which would prove whether or not a quantized quadrupole moment is present. Such higher-order-bulk responses could be measured in electronic systems, in which higher-order insulators were demonstrated earlier this year⁴. Future work could also extend the theoretical formalism to general external fields, rather than solely electromagnetic fields.

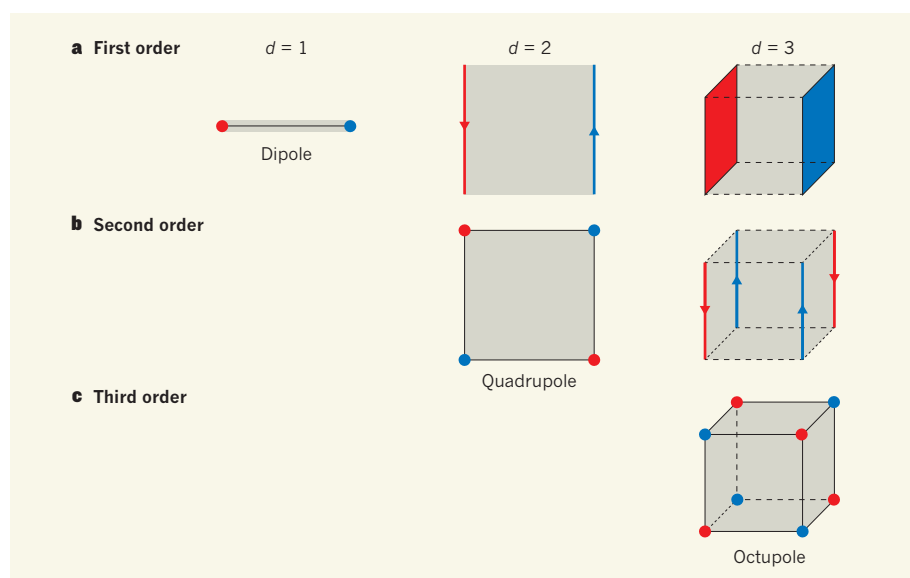


Figure 1 | Types of topological insulator. Materials known as topological insulators consist of a d -dimensional interior (grey) whose boundaries can host oscillations called topologically protected modes (red and blue; the two colours correspond to opposite electric charges). **a**, In first-order topological insulators, modes are localized in dimension $d=1$ (dipole topological insulators), travel along one-dimensional channels in $d=2$ and exist on surfaces in $d=3$. **b**, Three papers¹⁻³ report evidence for second-order topological insulators in $d=2$ (quadrupole systems), for which modes are localized to corners. Second-order topological insulators do not exist in $d=1$, and modes are supported along one-dimensional hinges in $d=3$. **c**, At third order, the minimum dimension is $d=3$, for which modes exist on corners (octupole topological insulators).

In terms of potential applications, it is not yet clear whether higher-order topological modes localized to corners or hinges have practical advantages over their conventional counterparts. For instance, higher-order topological insulators rely on the existence of crystal symmetries that typically limit the robustness of the edge modes. Moreover, it has been shown that protected modes can also be localized to points or lines of dimensionality lower than $(d-1)$ in ordinary topological insulators that have material defects^{11–14}.

Finally, one can speculate about such systems beyond third order — in other words, beyond the octupole moment. However, these are difficult to realize because of the unfortunate lack of spatial dimensions in our everyday world. Possible ways of overcoming this difficulty include resorting to ‘synthetic’ dimensions provided by internal degrees of freedom (such as the oscillation modes of a resonator), or artificially enhancing the connectivity of crystal lattices using long-range links¹⁵.

The authors’ experimental evidence for higher-order topological insulators illustrates the rapid transition from theoretical proposals to experimental realizations in current research on topological materials. We expect the next few years will be the time for such materials to prove their engineering worth. ■

Michel Fruchart is at the Lorentz Institute for Theoretical Physics, University of Leiden, 2333 CA Leiden, the Netherlands. **Vincenzo Vitelli** is at the James Frank Institute and in the Department of Physics, University of Chicago, Chicago, Illinois 60637, USA. e-mails: fruchart@lorentz.leidenuniv.nl; vitelli@uchicago.edu

1. Serra-Garcia, M. *et al.* *Nature* **555**, 342–345 (2018).
2. Peterson, C. W., Benalcazar, W. A., Hughes, T. L. & Bahl, G. *Nature* **555**, 346–350 (2018).
3. Imhof, S. *et al.* Preprint at <https://arxiv.org/abs/1708.03647> (2017).
4. Schindler, F. *et al.* Preprint at <https://arxiv.org/abs/1802.02585> (2018).
5. Benalcazar, W. A., Bernevig, B. A. & Hughes, T. L. *Science* **357**, 61–66 (2017).
6. Benalcazar, W. A., Bernevig, B. A. & Hughes, T. L. *Phys. Rev. B* **96**, 245115 (2017).
7. Schindler, F. *et al.* Preprint at <https://arxiv.org/abs/1708.03636> (2017).
8. Langbehn, J., Peng, Y., Trifunovic, L., von Oppen, F. & Brouwer, P. W. *Phys. Rev. Lett.* **119**, 246401 (2017).
9. Song, Z., Fang, Z. & Fang, C. *Phys. Rev. Lett.* **119**, 246402 (2017).
10. Bertoldi, K., Vitelli, V., Christensen, J. & van Hecke, M. *Nature Rev. Mater.* **2**, 17066 (2017).
11. Ran, Y., Zhang, Y. & Vishwanath, A. *Nature Phys.* **5**, 298–303 (2009).
12. Teo, J. C. Y. & Kane, C. L. *Phys. Rev. B* **82**, 115120 (2010).
13. Paulose, J., Chan, B. G. & Vitelli, V. *Nature Phys.* **11**, 153–156 (2015).
14. Baardink, G., Souslov, A., Paulose, J. & Vitelli, V. *Proc. Natl Acad. Sci. USA* **115**, 489–494 (2018).
15. Ozawa, T. *et al.* Preprint at <https://arxiv.org/abs/1802.04173> (2018).

IMMUNOLOGY

Melanin triggers antifungal defences

Melanins are enigmatic pigments that have many roles, and the melanin in pathogenic fungi can aid host infection. Identification of a mammalian protein that recognizes melanin now reveals an antifungal defence pathway. SEE LETTER P.382

ARTURO CASADEVALL

Most organisms produce numerous varieties of the highly diverse dark pigments known as melanins, which are among the last remaining biological frontiers with the unknown. These polymer molecules can act in protective or harmful ways, in biological functions as diverse as providing protection against DNA-damaging ultraviolet radiation¹ to bolstering fungal cell-wall strength². Melanins bolster microbial virulence³, including that of many disease-causing fungi. The presence of melanin can trigger an immune response in the infected organism⁴, but how this occurs was unknown. On page 382, Stappers *et al.*⁵ report the identification of a protein that can recognize a type of melanin produced by the fungus *Aspergillus fumigatus*. Their finding illuminates

the immune-system response to a fungal infection that can be lethal in people who have a suppressed immune system, such as those who have undergone transplantation surgery⁶.

Melanin pigments are stable free radicals, and, in animals and fungi, they are produced in membrane-bound organelles known as melanosomes, which shield the cell cytoplasm from the potentially damaging free-radical reaction needed for melanin production. They are insoluble and resistant to degradation by acids. These striking characteristics probably explain why their structures are difficult to analyse and are not fully understood. Host immune cells can trigger potentially damaging cell-signalling pathways in fungi. But such attacks can be neutralized by fungal melanin, which also reduces susceptibility to antifungal drugs³.

Human disease caused by fungi of the genus *Aspergillus* is called aspergillosis. If a



50 Years Ago

Like many other museums of its type, the Museum of Comparative Zoology has teaching and curatorial responsibilities, expeditions are organized to build up collections, and staff travel to study collections in other museums. Research conducted in the museum covers a wide range of topics — evolution, behaviour, ecology, zoogeography, physiology and biochemistry and taxonomy. Almost all the research produces results of interest to the evolutionist. One interesting find during the year was the discovery of a fossil insect from Cretaceous amber from New Jersey. This is the oldest known ant and is apparently virtually a missing link between ants and wasps. The presence of worker characteristics in these insects is evidence of the existence of social Hymenoptera as far back as about 100 million years. **From *Nature* 16 March 1968**

100 Years Ago

An announcement in the daily Press states that whale-meat furnished the principal article of food at a luncheon given in New York by the American Museum of Natural History to demonstrate the possibilities of whale-meat for home consumption, in order that the beef thus saved might be sent by America to relieve the scarcity prevailing among the Allies in Europe ... Unfortunately, we can do little to assist in this saving, for the whales in our home-waters cannot be ‘fished’, since neither ships nor men are available for the purpose ... It is to be hoped, however, that the fullest possible use will be made of the carcasses of the various species of Cetacea stranded around our coasts. Of course, no great quantity of meat would thus be obtained, but locally it should form a very welcome addition to the scanty meat rations now of necessity prevailing. **From *Nature* 14 March 1918**

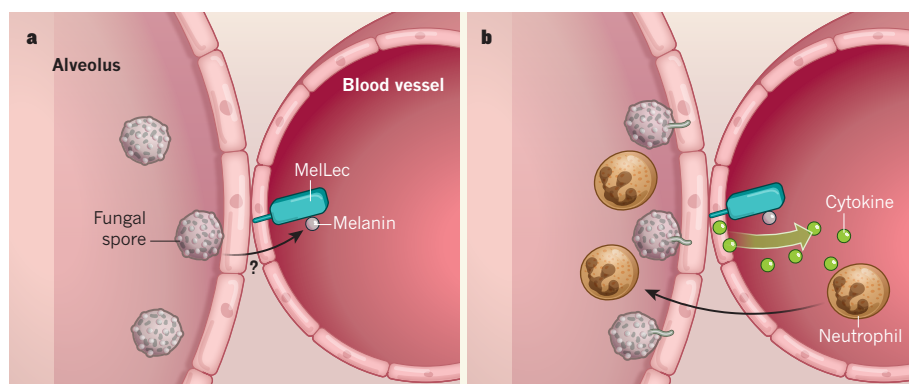


Figure 1 | A receptor that recognizes melanin and triggers antifungal defences. **a**, Infection by the fungus *Aspergillus fumigatus* can be lethal to certain people with weakened immune systems⁶, and many aspects of the body's immune response to such infections are unknown. *A. fumigatus* spores contain melanin and can infect air-filled sacs in the lungs called alveoli. Using mice, Stappers *et al.*⁵ investigated a protein family linked to antifungal defences⁸ called C-lectins, and found evidence that one of these proteins, named MelLec by the authors, can bind a type of melanin pigment present in fungi. Melanin pigment can be sensed by MelLec in cells lining blood vessels, but how melanin reaches MelLec is unknown. **b**, As infection progresses, the germinating spores form cellular protrusions. Melanin recognition by MelLec triggers the synthesis of cytokine molecules that can attract immune cells called neutrophils. Neutrophils can then enter the alveolus and target the infection.

lung infection takes hold in someone who has inhaled *A. fumigatus* spores, it can result in an infection that spreads elsewhere in the body. When *A. fumigatus* infects the lungs, host cells can trigger a cellular-degradation pathway called autophagy that aids fungal destruction. However, fungal melanin can inhibit autophagy⁷. Moreover, melanin is linked to inflammation.

The ability of melanin to target host defences, and the molecule's role in fungal virulence, raises the question of whether mammalian cells can recognize melanin. Stappers and colleagues investigated this by studying members of the C-type lectin protein family, which has previously been identified⁸ as being involved in antifungal defence. Using an *in vitro* biochemical approach, the authors tested whether any C-type lectins from mice can bind fungal spores from *A. fumigatus*. One of the proteins they tested could do so, and they named it MelLec.

The authors tested strains of *A. fumigatus* containing mutations that block steps in the melanin-synthesis pathway, and found that MelLec recognizes 1,8-dihydroxynaphthalene melanin. MelLec did not recognize other tested forms of melanin that are associated with fungal disease.

The authors found that mouse MelLec is expressed in the endothelial cells that line the surface of vessels forming the circulatory system. This suggests that it responds to infection after *A. fumigatus* has breached the lung defences in air-filled sacs called alveoli and moved farther into the body to reach the circulatory system (Fig. 1). In humans, MelLec is expressed in endothelial cells and in a type of immune cell known as a myeloid cell⁹.

The authors genetically engineered mice that lacked MelLec. These mice seemed normal,

but after treatment with molecules to induce immunosuppression and the injection of *A. fumigatus* spores into their bloodstream, they were more susceptible to infection than wild-type counterparts that had undergone the same treatment. Direct introduction of *A. fumigatus* into the lungs of mice lacking MelLec resulted in fewer immune cells called neutrophils entering the animals' lungs than was the case for wild-type mice, suggesting that melanin recognition by MelLec aids neutrophil recruitment to sites of infection. The authors found that the reduced neutrophil recruitment in mice lacking MelLec was linked to lower expression of neutrophil-attracting molecules called cytokines.

Although *A. fumigatus* is ubiquitous in the environment, not everyone with impaired immunity develops aspergillosis, suggesting that some individuals might be particularly vulnerable to the infection. To investigate this, Stappers and colleagues studied people who were in an immunosuppressed state following transplantation. Those who had a mutant version of MelLec in which a specific glycine amino-acid residue was replaced by alanine were more susceptible to infection by *A. fumigatus* than those who had the normal version of the protein. *In vitro* analysis of human cells revealed that this mutation is associated with decreased cytokine production in response to fungal exposure compared with cytokine production in cells containing the normal version of MelLec.

The identification of a MelLec mutation linked to susceptibility to fungal infection suggests an immediate clinical application in identifying patients at high risk of developing *Aspergillus* infections and who might benefit the most from antifungal treatments. Moreover, individuals with a functioning immune system can develop a hypersensitive reaction

to *Aspergillus*, a condition known as allergic pulmonary aspergillosis, and other MelLec mutations might be responsible for this predisposition.

As with all good scientific studies, the answer to the question of whether the body can sense melanin raises many additional questions. For example, how does melanin pigment on spores in the alveoli reach MelLec on cells located more internally? Perhaps when spores germinate and form cellular protrusions, these damage alveolar integrity and enable the spores to reach blood vessels. Another possibility is that spores are ingested and transported by macrophage cells of the immune system¹⁰.

Stapper and colleagues' work might mark the beginning of an era in which additional melanin-binding molecules are discovered. L-Dopa melanin and other types of melanin are pro-inflammatory⁴, so it seems reasonable to speculate that they are recognized by as-yet-unknown host proteins. Furthermore, MelLec offers a target for drug development because drugs that enhance its activity might boost immune responses to *Aspergillus* infection.

Like the discovery of the Toll-like receptor proteins that sense microbial infection in the fruit fly *Drosophila melanogaster*, Stapper and colleagues' identification of this first known melanin receptor arose from fungal-infection studies in model animals. At a time when researchers are increasingly urged to focus on studies with immediate clinical relevance, it is important to remember that transformative work often begins with model systems. Given that fungi are major pathogens targeting invertebrates, perhaps MelLec homologues exist in animal models such as *D. melanogaster* and the worm *Caenorhabditis elegans*, opening the door to the use of these organisms for additional investigation of this phenomenon. These and other studies building on the work of Stapper and colleagues might further our understanding of host-defence mechanisms. ■

Arturo Casadevall is in the Department of Molecular Microbiology and Immunology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland 21205, USA. e-mail: acasade1@jhu.edu

- Wang, Y. & Casadevall, A. *Appl. Environ. Microbiol.* **60**, 3864–3866 (1994).
- Brush, L. & Money, N. P. *Fungal Genet. Biol.* **28**, 190–200 (1999).
- Nosanchuk, J. D. & Casadevall, A. *Antimicrob. Agents Chemother.* **50**, 3519–3528 (2006).
- Mednick, A. J., Nosanchuk, J. D. & Casadevall, A. *Infect. Immun.* **73**, 2012–2019 (2005).
- Stappers, M. H. T. *et al. Nature* **555**, 382–386 (2018).
- Upton, A., Kirby, K. A., Carpenter, P., Boeckh, M. & Marr, K. A. *Clin. Infect. Dis.* **44**, 531–540 (2007).
- Chamilos, G. *et al. Autophagy* **12**, 888–889 (2016).
- Hardison, S. E. & Brown, G. D. *Nature Immunol.* **13**, 817–822 (2012).
- Sobanov, Y. *et al. Eur. J. Immunol.* **31**, 3493–3503 (2001).
- Shah, A. *et al. Am. J. Respir. Crit. Care Med.* **194**, 1127–1139 (2016).

This article was published online on 28 February 2018.

The landscape of genomic alterations across childhood cancers

A list of authors and affiliations appears at the end of the paper.

Pan-cancer analyses that examine commonalities and differences among various cancer types have emerged as a powerful way to obtain novel insights into cancer biology. Here we present a comprehensive analysis of genetic alterations in a pan-cancer cohort including 961 tumours from children, adolescents, and young adults, comprising 24 distinct molecular types of cancer. Using a standardized workflow, we identified marked differences in terms of mutation frequency and significantly mutated genes in comparison to previously analysed adult cancers. Genetic alterations in 149 putative cancer driver genes separate the tumours into two classes: small mutation and structural/copy-number variant (correlating with germline variants). Structural variants, hyperdiploidy, and chromothripsis are linked to *TP53* mutation status and mutational signatures. Our data suggest that 7–8% of the children in this cohort carry an unambiguous predisposing germline variant and that nearly 50% of paediatric neoplasms harbour a potentially druggable event, which is highly relevant for the design of future clinical trials.

Cure rates for childhood cancers have increased to about 80% in recent decades, but cancer is still the leading cause of death by disease in the developed world among children over one year of age^{1,2}. Furthermore, many children who survive cancer suffer from long-term sequelae of surgery, cytotoxic chemotherapy, and radiotherapy, including mental disabilities, organ toxicities, and secondary cancers³. A crucial step in developing more specific and less damaging therapies is the unravelling of the complete genetic repertoire of paediatric malignancies, which differ from adult malignancies in terms of their histopathological entities and molecular subtypes⁴. Over the past few years, many entity-specific sequencing efforts have been launched, but the few paediatric pan-cancer studies thus far have focused only on mutation frequencies, germline predisposition, and alterations in epigenetic regulators^{4–6}.

We have carried out a broad exploration of cancers in children, adolescents, and young adults, by incorporating small mutations and copy-number or structural variants on somatic and germline levels, and by identifying putative cancer genes and comparing them to those previously reported in adult cancers by The Cancer Genome Atlas (TCGA)⁷. We have also examined mutational signatures and potential drug targets. The compendium of genetic alterations presented here is available to the scientific community at <http://www.pedpancan.com>.

This integrative analysis includes 24 types of cancer and covers all major childhood cancer entities, many of which occur exclusively in children⁸ (Fig. 1, Supplementary Table 1). Ninety-five per cent of the patients in this study were diagnosed during childhood or adolescence (aged 18 years or younger) and 5% as young adults (up to 25 years) (Extended Data Fig. 1a). This study is biased towards central nervous system tumours, and is complemented by an additional study of a non-overlapping paediatric cohort with mainly leukaemias and extra-cranial solid tumours⁹.

We compiled paired-end Illumina-based sequencing data for 961 tumours (914 individual patients) from previous cancer-type specific studies (see Methods and Supplementary Note 1) including 547 whole-genome sequences (WGS, median coverage 37×) and 414 whole-exome sequences (WES, 121×) partially complemented by low-coverage whole genomes (Supplementary Tables 1, 2). Tumour and matched germline samples were processed with standardized pipelines to detect single nucleotide variants (SNVs), short insertions and

deletions (indels), copy-number variants (CNVs) and other structural variants. Secondary (relapse) tumours ($n = 82$, including 47 matched to primaries) were analysed separately from the main primary cohort ($n = 879$).

Mutation frequencies across cancer types

Coding somatic SNV (93%) and indel (7%) counts correlated across all samples ($n = 879$) ($R = 0.27$, $P = 9.1 \times 10^{-5}$; Extended Data Fig. 1b, c). Mutation frequencies varied between cancer types (0.02–0.49 mutations per Mb) and were overall 14 times lower than in adult cancers⁷ (0.13 versus 1.8 mutations per Mb, TCGA data; Fig. 1, Extended Data Fig. 1c, Supplementary Table 3). Relapse tumours harboured significantly more mutations than primary tumours ($P = 0.0015$, excluding highly mutated tumours; Extended Data Fig. 1d).

Tumours with more than 10 mutations per Mb have been referred to as ‘hypermutators’, and are often related to deficiencies in mismatch repair (MMR)^{10,11}. In this cohort, hypermutation occurred exclusively in H3.3 or H3.1 K27-wildtype (K27wt) high-grade gliomas with biallelic germline mutations in *MSH6* or *PMS2*, with an extremely high mutational burden similar to the highest among adult tumours (in *POLE*- or *POLQ*-mutated carcinomas)^{7,12} (Fig. 1). Some paediatric tumours had a mutational burden below this threshold, but markedly above average (2–10 mutations per Mb, referred to as ‘paediatric highly mutated’), including several K27wt high-grade gliomas with monoallelic germline variants in *MSH2*, *MSH6* or *PMS2* (Fig. 1). Whether these highly mutated tumours respond to immune checkpoint inhibitors, as described for paediatric glioblastoma, should be of clinical interest¹³.

As in previous reports, the somatic mutation burden increased with patient age ($R = 0.39$, $P = 2.9 \times 10^{-6}$), except in Burkitt’s lymphoma (immunoglobulin hypermutation) and tumours with ‘kataegis’ events of localized hypermutation at double-stranded breakpoints^{14,15} (Extended Data Fig. 1e, f). Both SNVs ($R = 0.37$, $P = 1.0 \times 10^{-5}$) and indels ($R = 0.27$, $P = 5.4 \times 10^{-4}$) correlated with patient age overall, although within some cancers (for example, acute lymphoblastic leukaemia (ALL), Ewing’s sarcoma, and rhabdomyosarcoma), we observed almost random mutational loads ($R < 0.2$). Rhabdomyosarcomas were largely dominated by embryonal tumours with more mutations than the few alveolar cases (median 0.27 versus 0.12 mutations per Mb, $P = 0.002$).

Mutational processes in childhood cancers

Most cancer types predominantly harboured C > T transitions ($\geq 30\%$ of SNVs in two-thirds of cancer types) linked to mutational signature 1, whose previously described age-association occurred in some paediatric brain tumours^{15,16} ($P < 0.05$; Extended Data Figs 1g, 2a–c). Mutational signatures, possibly reflecting biochemical cellular processes, have previously been investigated for many, mainly adult, cancers¹⁵. In this paediatric cohort (WGS, $n = 503$), we found evidence for major contributions of 16 out of 30 published signatures and also identified one new signature¹⁵ (Fig. 2, Extended Data Fig. 2a, Supplementary Table 4). This ‘signature P1’, which is distinct from any previously documented signatures and harbours elevated C > T mutations in a CCC/CCT context, occurred in several atypical teratoid rhabdoid tumours (ATRTs) and one ependymoma (Fig. 2, Extended Data Fig. 2d, Supplementary Table 5). Its activity correlated with ‘multiple nucleotide variants’ (MNVs; $R = 0.87$, $P = 1.1 \times 10^{-12}$), but no particular loci or genes were mutually altered in the affected tumours (Extended Data Fig. 2d). Notably, all ATRTs with signature P1 were in the recently defined subgroup ‘SHH’, and even within one proposed methylation subset of these¹⁷ ($P = 0.003$, Wilcoxon rank-sum test; Extended Data Fig. 2d). Signatures 16 and 18 were heterogeneously represented within several cancer types, with signature 16 being most prominent in pilocytic astrocytomas, and signature 18, previously proposed to be associated with oxidative DNA damage and related to C > A transversions, in neuroblastomas, rhabdomyosarcomas, and other tumours with multiple structural variants^{15,18} (Extended Data Figs 1g, 2a, c, 3a).

Signature 3, the ‘canonical’ double-stranded break signature linked to mutations in *BRCA1* or *BRCA2* or to a ‘BRCAness’ phenotype, and signatures 8 (recently linked to *BRCA2* or *PALB2* germline mutations in medulloblastomas; S. M. Waszak *et al.*, personal communication) and 13 were linked to chromothripsis and *TP53* mutations. This was particularly true for *TP53* germline-mutated SHH medulloblastomas,

and similarly for adrenocortical carcinomas and rhabdomyosarcomas (Extended Data Fig. 3b, c). Overall, signatures 3, 8, and 13 were more pronounced in cancer types with higher genomic instability (that is, structural variants; Extended Data Fig. 2e).

Germline variants in cancer predisposition genes

A recent study of more than 1,000 patients estimated that about 8% of children with cancer harbour a hereditary predisposition⁵. Accordingly, in our cohort ($n = 914$ individual patients, about 25% of samples overlapping with the previous study), 7.6% of samples were determined as being likely to be associated with a pathogenic germline variant^{5,19} (162 genes investigated; Supplementary Tables 6, 7). No general age-of-onset bias was observed in patients with a predisposition; however, onset was later in germline MMR-deficient patients ($P = 0.0001$), even within the high-grade glioma sub-cohort ($P = 0.001$).

Hereditary predisposition was most common in adrenocortical carcinomas (50%) and hypodiploid B-ALL (28%), followed by K27wt high-grade gliomas, ATRTs, SHH medulloblastomas, and retinoblastomas (15–25% each; Fig. 3a). Compared to the previous study, *LZTR1*, *TSC2*, and *CHEK2* emerged as new putative predisposition genes, and possible new associations, such as *SDHA* with medulloblastoma, were detected⁵ (Fig. 3b).

Most germline variants were related to DNA repair genes from mismatch (*MSH2*, *MSH6*, *PMS2*) and double-stranded break (*TP53*, *BRCA2*, *CHEK2*) repair (Fig. 3b, c). Both groups are clinically relevant: patients with constitutional MMR deficiency could be candidates for immune checkpoint inhibition¹³ (Figs 1, 3b, c). Carriers of *TP53* germline mutations (Li–Fraumeni syndrome), here most common in adrenocortical carcinomas, hypodiploid B-ALL, SHH medulloblastomas, and K27wt high-grade gliomas, are at a 50% risk for early-onset cancer compared to 1% overall, and are susceptible to treatment-induced secondary oncogenesis^{2,20–22} (Fig. 3b). Correcting the predisposition frequency of 7.6% in this cohort for the relative incidence of

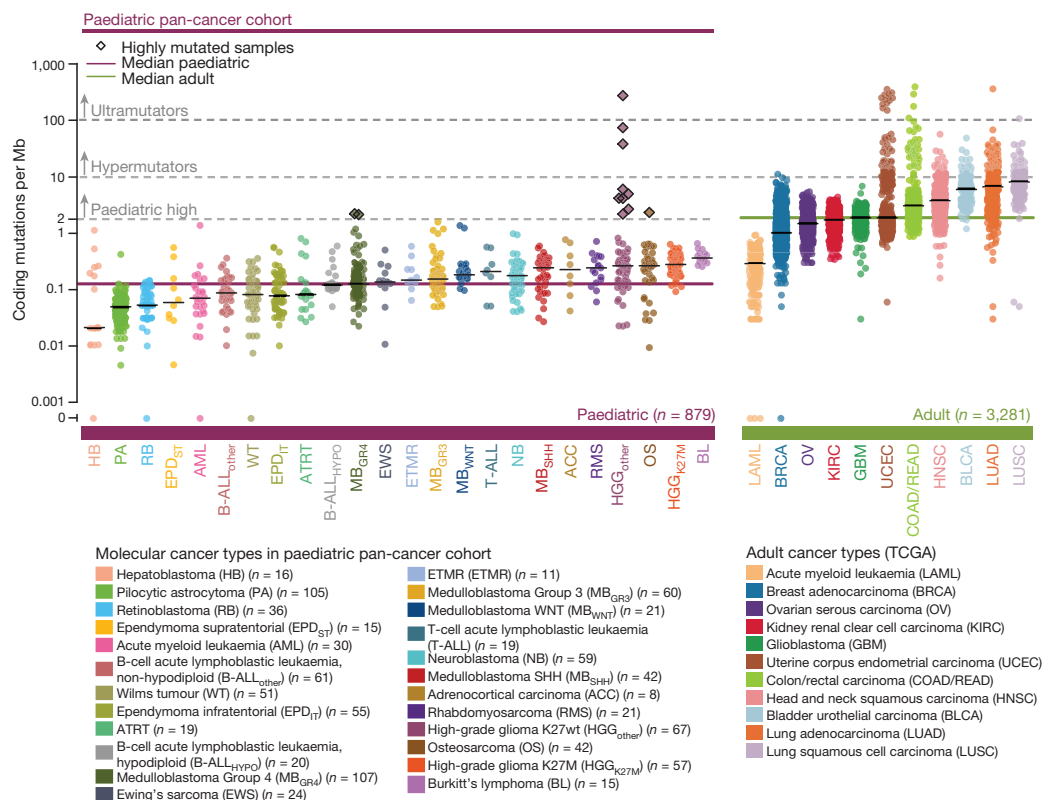


Figure 1 | Somatic mutations in the paediatric pan-cancer cohort. Somatic coding mutation frequencies in 24 paediatric ($n = 879$) primary tumours and 11 adult ($n = 3,281$) cancer types (TCGA). Hypermutated

and highly mutated samples are separated by dashed grey lines and highlighted with black squares. Median mutation loads are shown as solid lines (black, cancer types; purple, all paediatric; green, all adult).

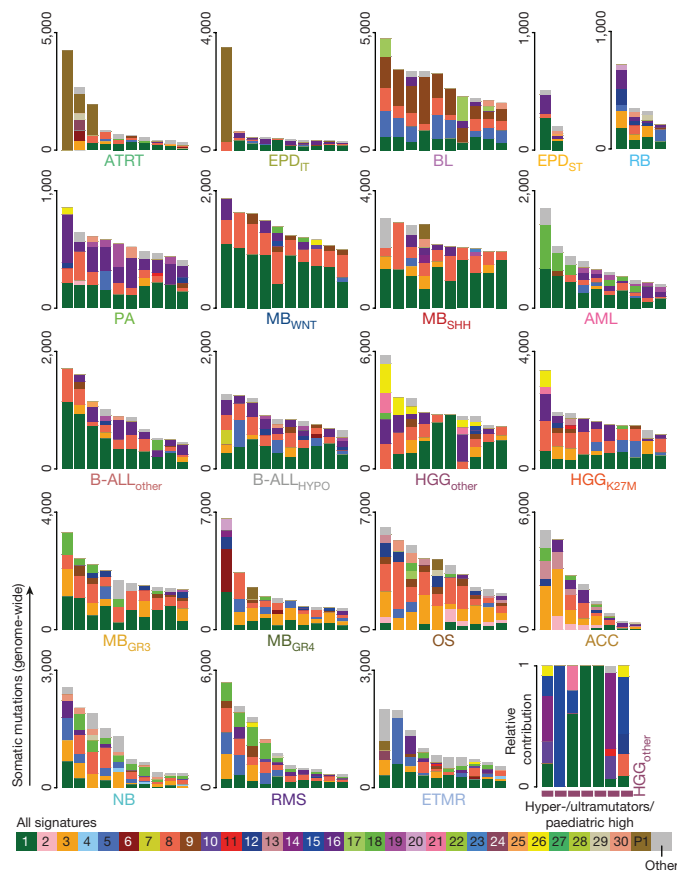


Figure 2 | Mutational processes active in paediatric cancers.

Contributions of thirty known and one novel mutational signature to the somatic mutations for the ten most frequently mutated samples per cancer type; each bar represents one individual tumour.

cancer types as a whole, we find that approximately 6% of all childhood cancer patients may carry a causative germline variant (Fig. 3d).

Significance analysis identifies cancer driver genes

Genome-wide analysis for significant mutation clusters ($n = 538$, WGS excluding hypermutators) identified non-coding mutations in the *TERT* promoter in 2.5% of tumours (Extended Data Fig. 4a, b, Supplementary Table 8). Further high-confidence clusters corresponded to coding mutations in frequently mutated genes (*TP53*, *H3F3A*, *CTNNB1*), and to localized hypermutation at the rearranged *MYC* locus in Burkitt's lymphoma, while the bulk were classified as likely technical artefacts²³ (Extended Data Fig. 4b).

MuSiC identified 77 significantly mutated genes (SMGs), which were ranked according to their pan-cancer mutation frequency²⁴ (Fig. 4, Supplementary Tables 9, 10). Most SMGs were mutually exclusively mutated across cancer types, demonstrating specificity of single putative driver genes in childhood cancers as compared to more frequent co-mutation in adult cancers in the TCGA study⁷ (Extended Data Fig. 4c–e). None of the SMGs showed a bias towards samples with higher mutation frequencies. The allele frequencies of mutations in SMGs were higher than in non-SMGs, and ranked higher in individual tumours, suggesting an early clonal occurrence of these likely driver events (Extended Data Fig. 4f). Two additional SMGs emerged from analysis of the relapse tumours ($n = 82$): *PRPS1* and *NT5C2*, both of which have been previously implicated in disease progression and chemotherapy resistance^{25,26} (Extended Data Fig. 4g).

Genes linked to epigenetic modification emerged as the most common (25% of tumours, 23 of 24 cancer types) and the largest (20%) group of SMGs (Extended Data Fig. 5a). Compared to a previous study⁶, for example, we also detected *ARID1A* and *BCOR*. Transcriptional

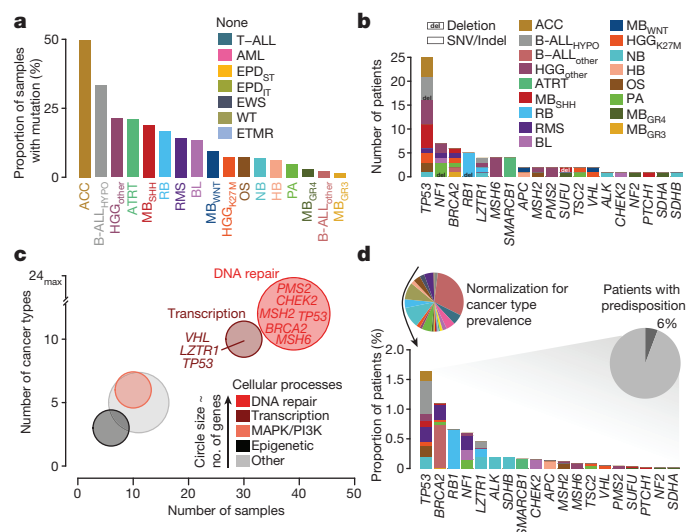


Figure 3 | Germline mutations in cancer predisposition genes.

a, Frequency of patients with a pathogenic germline mutation per cancer type ($n = 914$ tumours). **b**, Mutated genes sorted by number of affected samples (del, copy-number alterations; others, SNVs/indels). **c**, Cellular processes associated with cancer predisposition genes. **d**, Frequency of germline mutations adjusted for incidence and estimated total proportion of childhood cancers likely to be linked to hereditary predisposition.

regulators and MAP-kinase-associated genes accounted for 12–15% of SMGs. *TP53* was the only DNA repair gene among somatic SMGs, in contrast to the multiple DNA repair-related germline mutations, and also in contrast to adult cancers (9% of SMGs, TCGA)⁷. PI3K-associated SMGs are the most commonly altered (31%) genes in adult cancers, compared to only 3% in paediatric cancers, which could be related to their often late occurrence in the evolution of multi-hit adult cancers²⁷ (Extended Data Fig. 5a).

Forty-seven per cent of paediatric tumours harboured at least one SMG mutation, with most tumours (57%) having only one. SMG mutations were rare (<15%) in ependymomas, hepatoblastomas, Ewing's sarcomas (driven by *EWSR1* fusions instead of by point mutations²⁸), and pilocytic astrocytomas, and common (>90%) in K27M high-grade gliomas, WNT medulloblastomas, and Burkitt's lymphomas. By contrast, 93% of adult cancers harbour at least one mutation in an (adult cancer-related) SMG and 76% in multiple SMGs⁷ (Extended Data Fig. 5b). In line with the accompanying paediatric pan-cancer study⁹, only around 30% of paediatric SMGs overlapped with adult SMGs (Extended Data Fig. 5c). On the basis of incidence-normalized mutation frequencies, *TP53* is predicted to be the most common somatically mutated gene (4% of childhood tumours), followed by *KRAS*, *ATRX*, *NF1*, and *RB1* (1–2% of tumours); in adult cancers, with similarly normalized data, *TP53* is also the most commonly mutated gene, albeit ten times more frequently (Extended Data Fig. 5d).

Assessment of high functional impact mutations (OncodriveFM)²⁹ revealed well-known tumour suppressor genes (TSGs) such as *TP53*, *ATRX*, *SMARCA4*, and *RB1*, and further putative TSGs, including *FMRI* in SHH/WNT medulloblastomas and *MALRD1* (also known as *C10orf112*) in rhabdomyosarcomas (Extended Data Fig. 6a). Locally clustered 'hotspot mutations' (OncodriveClust)^{29,30} identified known oncogenes, such as *CTNNB1*, *PIK3CA*, *KRAS*, and *BRAF*, proposed oncogenes (*ACVR1*, *KBTBD4*, *TBR1*), and possible new candidates, such as *SF3B1*, in Group 4 medulloblastomas (Extended Data Fig. 6b).

Recurrent structural and copy-number variants

The degree of genomic instability (that is, the number of structural variants, including insertions, deletions, translocations, and inversions), varied substantially (median 1–434 structural variants) across

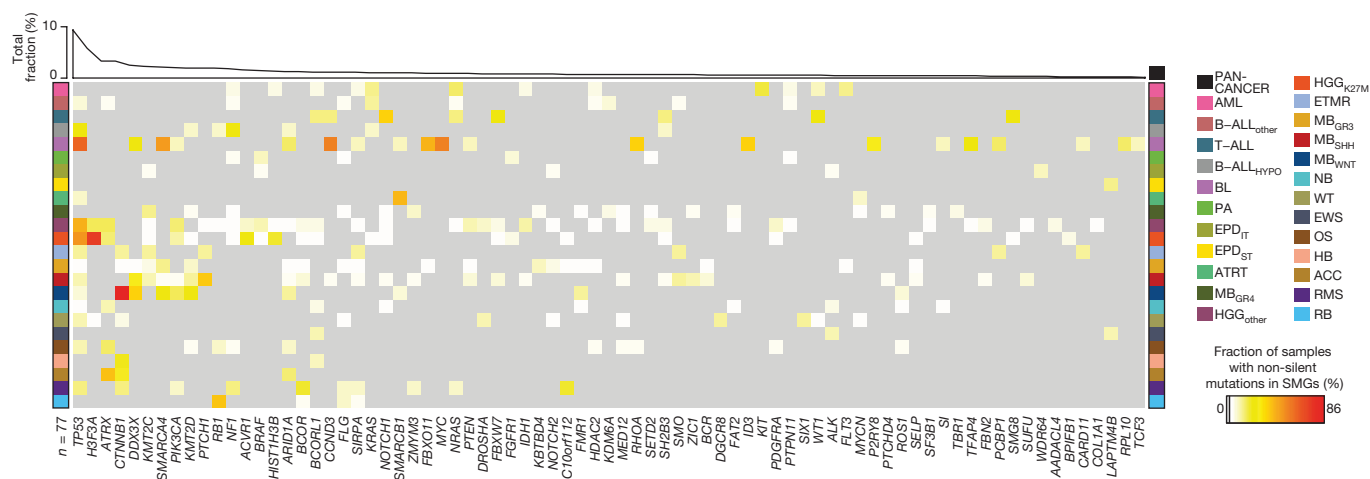


Figure 4 | Significantly mutated genes in paediatric compared to adult cancer types. Percentage of tumours with non-silent mutations in 77 SMGs for 24 paediatric tumour types ($n=879$ tumours) and the pan-cancer cohort.

cancer types (WGS, $n = 539$), with more than 1,000 structural variants in individual samples of adrenocortical carcinoma and osteosarcoma (Fig. 5a, Supplementary Table 11). Genomic instability correlated with germline ($P = 3 \times 10^{-15}$) and somatic ($P = 2 \times 10^{-4}$) *TP53* mutations across all samples, but differed markedly between cancer types—again suggesting cancer type-specific effects of DNA repair (Fig. 5b, Extended Data Figs 3b, 7a).

Genomically unstable cancers were also more often hyperdiploid³¹ (Supplementary Table 12). Twelve per cent of tumours had a ploidy of four or more, 72% retained a near-diploid state (ploidy 1.5–2.5), and hypodiploidy was observed mainly in hypodiploid B-ALLs (Extended Data Fig. 7b). Hyperdiploidy was associated with somatic ($P = 0.005$) and germline ($P = 0.003$) *TP53* mutations, in line with a role for mutant *TP53* in the bypassing of the G1 tetraploidy checkpoint³² (Extended Data Fig. 7c–e). Chromothripsis was also often observed in hyperdiploid cancers and co-occurred with somatic ($P = 2.3 \times 10^{-10}$) and germline *TP53* ($P = 5 \times 10^{-8}$) mutations in 50% and 66% of these tumours, compared to 8% in *TP53* wild-type tumours^{33–35} (Extended Data Fig. 7f–h, Supplementary Table 13).

Thirty-four regions recurrently altered by copy-number changes (17 amplified, 17 deleted) were identified using GISTIC2.0 (WGS, $n = 516$)³⁶; candidate driver genes were assigned to each based on known cancer genes and literature review (Fig. 5c, Extended Data Fig. 8a, b, Supplementary Tables 14–17). Alterations per cancer type are summarized in Extended Data Fig. 9.

Recurrently amplified regions contained known oncogenes, including *MYC*, *MYCN*, or *GLI2*, with 11 regions involving high-level amplifications (at least 5-fold gain) (Extended Data Fig. 8b). Further interesting regions included 17q11.2 with 61 genes, containing *NCOR1* as a potential candidate, and a region on 12q24.31 near (~0.1 Mb) the proposed oncogene *KDM2B*^{37,38}. Recurrently deleted regions were predominantly associated with epigenetic or cell cycle regulators, most commonly *TP53*, *PTEN*, *SETD2*, and *CDKN2A* or *CDKN2B*. Further potential tumour suppressors included *RAD51D* on 17q12 and *FOXF1* on 16q24.1, with significant loss across the cohort³⁹.

As evidenced by recurrent structural variation outside genes (based on breakpoint clusters in 10-kb windows), rearrangements linked to enhancer hijacking were also found, involving *GFI1B* and *DDX31* in medulloblastomas and *TERT* in neuroblastomas^{40,41}. Together with genes directly affected by breakpoints, in total 70 structural variant-related putative cancer genes were found, many associated with cell cycle or growth (for example, the tumour suppressor *PTPRD*) or epigenetic regulators (such as *SUZ12*)^{42,43} (Extended Data Fig. 8c, Supplementary Tables 18, 19). Cancer type-specific events that

occurred together with high expression (data derived from Northcott *et al.*⁴⁴) included alterations of *RIMS2*⁴⁵.

The analysed genomic alterations were combined into 166 ‘likely functional events’ (LFEs) affecting 149 genes, classified as M-(mutation)-type or as SC-(structural/copy-number variant)-type (Extended Data Fig. 10a, Supplementary Table 20). Along the ‘cancer genome hyperbola’, individual tumours (WGS, $n = 539$) differentiated between an M-class (more M-type LFEs) and an SC-class (more SC-type LFEs)⁴⁶ (Extended Data Fig. 10b, Supplementary Table 21). Fifty-five per cent of tumours were exclusive to one class, 27% were mixed but dominated by one type of LFE, 8% were ambiguous, and 10% had no LFEs (which may be of particular interest in assessing other tumour-driving events at the epigenetic or transcriptomic level). Germline MMR mutations were enriched in the M-class, and germline *TP53* mutations in the SC-class ($P = 0.0003$ and $P = 0.05$, respectively, Fisher’s exact test; Extended Data Fig. 10c). Individual cancer types displayed varying relative distributions of mutation classes (Extended Data Fig. 10d).

Drug targets in childhood cancers

To assess the status of druggability of childhood cancers, the cohort ($n = 675$ with full genomic information; WES-only, $n = 39$; see Methods) was screened for potentially druggable events¹⁹ (PDEs, that is, alterations in 179 genes with a directly or indirectly targeted treatment currently available or under development; Supplementary Table 22). This analysis revealed 453 PDEs in 59 genes, including 3% germline events (Supplementary Table 23). Most cancer types had tumours with PDEs related to both M- and SC-type (Fig. 6a). Most commonly, PDEs occurred in Burkitt's lymphomas and pilocytic astrocytomas, while none were detected in ependymomas or hepatoblastomas (although the latter lacked information regarding CNVs or structural variants). Associated pathways included RTK/MAPK signalling, transcriptional regulation, cell cycle control, and DNA repair (Fig. 6a).

When the data are normalized for relative cancer incidence, 52% of all primary paediatric tumours may harbour a PDE (Fig. 6b); this might be an underestimate, given that some structural variants may not have been detected by this approach (for example, the common *MYC* translocations in Burkitt's lymphoma)²³. After incidence adjustment, MAPK signalling and cell cycle control were most commonly affected. Notably, the PDEs often varied between primary and relapse tumours from one patient ($n = 41$): only 37% of primary tumours with PDEs retained these upon progression, while most of them partially or completely gained or lost events. This highlights the need for profiling of the current tumour when considering personalized therapy.

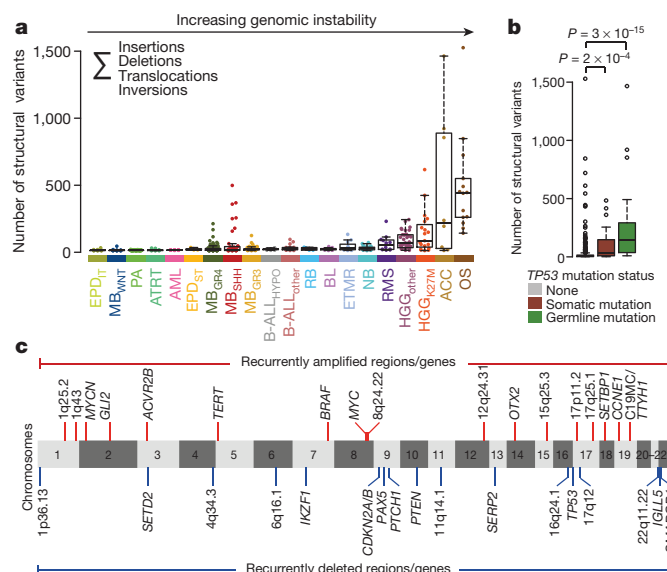


Figure 5 | Genomic instability and recurrent copy-number alterations. **a**, Frequency of structural variants (SVs) across cancer types ($n = 539$ tumours). **b**, Structural variant load from **a** across all tumours in relation to *TP53* mutations (generalized linear model, confidence interval 0.95). **a**, **b**, Quartiles, range of whiskers: $1.5 \times$ interquartile range. **c**, Genomic regions with significant copy-number changes (red, gains or amplifications; blue, deletions; $n = 516$ tumours).

Discussion

Our analysis of this pan-cancer compendium outlines the landscape of genomic alterations across multiple childhood cancer types. Although some alteration types and rarer entities are still under-represented and significance analyses are probably limited, this dataset of nearly 1,000 tumours (which can be explored at <http://www.pedpancan.com>) provides an unprecedented data resource for paediatric cancer research, further complemented by the accompanying pan-cancer study⁹ (<https://pecan.stjude.org/proteinpaint/study/pan-target>). The multiple differences found compared to previous studies of adult tumours emphasize the need to consider paediatric cancers separately, further demonstrating a need for mechanism-of-action driven drug development for paediatric indications⁴⁷.

The predicted frequency of pathogenic germline variants in 6% of patients, together with previous findings, demonstrates the relevance of genetic predisposition in childhood cancer⁵. Germline *TP53* variants, which are clinically highly important, are estimated for 1.5% of children with cancer, and for more than 10% within individual cancer types. Genetic counselling should thus be systematically considered, particularly for patients with indicated high-risk entities.

Although stratified targeted treatment is currently incorporated only rarely into first-line therapy for paediatric cancer patients, our finding that nearly 50% of primary childhood tumours harbour a potentially targetable genetic event is encouraging. It also highlights the need for personalized profiling for each patient, both to increase diagnostic accuracy and to exploit the potential for potentially more effective and less harmful precision therapies. This may also transcend the direct targeting of genes or pathways, for example, through immune checkpoint inhibition in hypermutated tumours¹³ or through PARP inhibition in genomically unstable ('BRCAness') tumours⁴⁸. It is hoped that ongoing personalized medicine approaches for patients at relapse will give initial information on the use and effectiveness of such targeted drugs (for example, in the clinical trials pedMATCH-NCT03155620; eSMART-NCT02813135; INFORM¹⁹). Additional longitudinal monitoring, for example using serial liquid biopsies, may further improve our understanding of tumour biology and the development of resistance

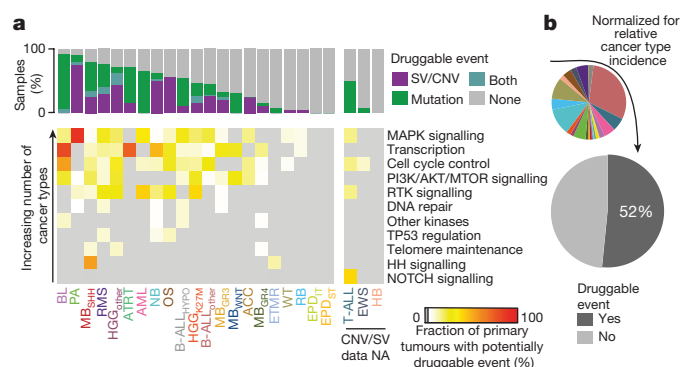


Figure 6 | Potentially druggable events in paediatric cancers.

a, Proportion of primary tumours with potentially druggable events and associated biological pathways, per cancer type ($n = 675$ tumours with complete genomic information). NA, not available. **b**, Proportion of patients with potentially druggable events, projected after normalization for incidence.

mechanisms, and shed light on therapeutic challenges such as tumour heterogeneity.

In summary, this multi-faceted pan-cancer analysis provides a valuable resource for assessing genomic alterations across the spectrum of paediatric tumours. While there are undoubtedly more discoveries to come in terms of expanded cohorts and whole-genome and transcriptome analysis, we believe that this study provides a strong basis for functional follow-up and investigation of potential therapeutic targets in this specific patient population.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 February; accepted 24 December 2017.

Published online 28 February 2018.

- Pui, C. H., Gajjar, A. J., Kane, J. R., Qaddoumi, I. A. & Pappo, A. S. Challenging issues in pediatric oncology. *Nat. Rev. Clin. Oncol.* **8**, 540–549 (2011).
- Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2016. *CA Cancer J. Clin.* **66**, 7–30 (2016).
- Kopp, L. M., Gupta, P., Pelayo-Katsanis, L., Wittman, B. & Katsanis, E. Late effects in adult survivors of pediatric cancer: a guide for the primary care physician. *Am. J. Med.* **125**, 636–641 (2012).
- Downing, J. R. et al. The Pediatric Cancer Genome Project. *Nat. Genet.* **44**, 619–622 (2012).
- Zhang, J. et al. Germline mutations in predisposition genes in pediatric cancer. *N. Engl. J. Med.* **373**, 2336–2346 (2015).
- Huether, R. et al. The landscape of somatic mutations in epigenetic regulators across 1,000 paediatric cancer genomes. *Nat. Commun.* **5**, 3630 (2014).
- Kandoth, C. et al. Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339 (2013).
- Howlader, N. et al. *SEER Cancer Statistics Review, 1975–2012*, National Cancer Institute (National Cancer Institute, SEER Program, NIH, 2014).
- Ma, X. et al. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature* <https://doi.org/10.1038/nature25795> (2018).
- Yuan, Y. et al. Assessing the clinical utility of cancer genomic and proteomic data across tumor types. *Nat. Biotechnol.* **32**, 644–652 (2014).
- Campbell, B. B. et al. Comprehensive analysis of hypermutation in human cancer. *Cell* **171**, 1042–1056 (2017).
- Kandoth, C. et al. Integrated genomic characterization of endometrial carcinoma. *Nature* **497**, 67–73 (2013).
- Bouffet, E. et al. Immune checkpoint inhibition for hypermutant glioblastoma multiforme resulting from germline biallelic mismatch repair deficiency. *J. Clin. Oncol.* **34**, 2206–2211 (2016).
- Miltholland, B., Auton, A., Suh, Y. & Vijg, J. Age-related somatic mutations in the cancer genome. *Oncotarget* **6**, 24627–24635 (2015).
- Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
- Alexandrov, L. B. et al. Clock-like mutational processes in human somatic cells. *Nat. Genet.* **47**, 1402–1407 (2015).
- Johann, P. D. et al. Atypical teratoid/rhabdoid tumors are comprised of three epigenetic subgroups with distinct enhancer landscapes. *Cancer Cell* **29**, 379–393 (2016).

18. Pilati, C. *et al.* Mutational signature analysis identifies MUTYH deficiency in colorectal cancers and adrenocortical carcinomas. *J. Pathol.* **242**, 10–15 (2017).
 19. Worst, B. C. *et al.* Next-generation personalised medicine for high-risk paediatric cancer patients—The INFORM pilot study. *Eur. J. Cancer* **65**, 91–101 (2016).
 20. Sorrell, A. D., Espenschied, C. R., Culver, J. O. & Weitzel, J. N. Tumor protein p53 (TP53) testing and Li-Fraumeni syndrome: current status of clinical applications and future directions. *Mol. Diagn. Ther.* **17**, 31–47 (2013).
 21. Kappel, S. *et al.* TP53 germline mutation may affect response to anticancer treatments: analysis of an intensively treated Li-Fraumeni family. *Breast Cancer Res. Treat.* **151**, 671–678 (2015).
 22. Heymann, S. *et al.* Radio-induced malignancies after breast cancer postoperative radiotherapy in patients with Li-Fraumeni syndrome. *Radiat. Oncol.* **5**, 104 (2010).
 23. Klapproth, K. & Wirth, T. Advances in the understanding of MYC-induced lymphomagenesis. *Br. J. Haematol.* **149**, 484–497 (2010).
 24. Dees, N. D. *et al.* MuSiC: identifying mutational significance in cancer genomes. *Genome Res.* **22**, 1589–1598 (2012).
 25. Mullighan, C. G. Mutant PRPS1: a new therapeutic target in relapsed acute lymphoblastic leukemia. *Nat. Med.* **21**, 553–554 (2015).
 26. Tzoneva, G. *et al.* Activating mutations in the NT5C2 nucleotidase gene drive chemotherapy resistance in relapsed ALL. *Nat. Med.* **19**, 368–371 (2013).
 27. Verlaet, W. *et al.* Somatic mutation in *PIK3CA* is a late event in cervical carcinogenesis. *J. Pathol. Clin. Res.* **1**, 207–211 (2015).
 28. Delattre, O. *et al.* Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. *Nature* **359**, 162–165 (1992).
 29. Gonzalez-Perez, A. *et al.* IntOGen-mutations identifies cancer drivers across tumor types. *Nat. Methods* **10**, 1081–1082 (2013).
 30. Tamborero, D. *et al.* Comprehensive identification of mutational cancer driver genes across 12 tumor types. *Sci. Rep.* **3**, 2650 (2013).
 31. Zack, T. I. *et al.* Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **45**, 1134–1140 (2013).
 32. Margolis, R. L., Lohez, O. D. & Andreassen, P. R. G1 tetraploidy checkpoint and the suppression of tumorigenesis. *J. Cell. Biochem.* **88**, 673–683 (2003).
 33. Mardin, B. R. *et al.* A cell-based model system links chromothripsis with hyperploidy. *Mol. Syst. Biol.* **11**, 828 (2015).
 34. Forment, J. V., Kaidi, A. & Jackson, S. P. Chromothripsis and cancer: causes and consequences of chromosome shattering. *Nat. Rev. Cancer* **12**, 663–670 (2012).
 35. Rausch, T. *et al.* Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. *Cell* **148**, 59–71 (2012).
 36. Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, R41 (2011).
 37. Merve, A. *et al.* Polycomb group gene BMI1 controls invasion of medulloblastoma cells and inhibits BMP-regulated cell adhesion. *Acta Neuropathol. Commun.* **2**, 10 (2014).
 38. He, J., Kallin, E. M., Tsukada, Y. & Zhang, Y. The H3K36 demethylase Jhdmlb/Kdm2b regulates cell proliferation and senescence through p15^{Ink4b}. *Nat. Struct. Mol. Biol.* **15**, 1169–1175 (2008).
 39. Tamura, M. *et al.* Forkhead transcription factor FOXF1 is a novel target gene of the p53 family and regulates cancer cell migration and invasiveness. *Oncogene* **33**, 4837–4846 (2014).
 40. Northcott, P. A. *et al.* Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature* **511**, 428–434 (2014).
 41. Valentijn, L. J. *et al.* TERT rearrangements are frequent in neuroblastoma and identify aggressive tumors. *Nat. Genet.* **47**, 1411–1414 (2015).
 42. Veeriah, S. *et al.* The tyrosine phosphatase PTPRD is a tumor suppressor that is frequently inactivated and mutated in glioblastoma and other human cancers. *Proc. Natl Acad. Sci. USA* **106**, 9435–9440 (2009).
 43. Cao, R. & Zhang, Y. SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. *Mol. Cell* **15**, 57–67 (2004).
 44. Northcott, P. A. *et al.* The whole-genome landscape of medulloblastoma subtypes. *Nature* **547**, 311–317 (2017).
 45. Kaeser, P. S., Deng, L., Fan, M. & Südhof, T. C. RIM genes differentially contribute to organizing presynaptic release sites. *Proc. Natl Acad. Sci. USA* **109**, 11830–11835 (2012).
 46. Ciriello, G. *et al.* Emerging landscape of oncogenic signatures across human cancers. *Nat. Genet.* **45**, 1127–1133 (2013).
 47. Pearson, A. D. *et al.* Implementation of mechanism of action biology-driven early drug development for children with cancer. *Eur. J. Cancer* **62**, 124–131 (2016).
 48. Cerrato, A., Morra, F. & Celetti, A. Use of poly ADP-ribose polymerase [PARP] inhibitors in cancer cells bearing DDR defects: the rationale for their inclusion in the clinic. *J. Exp. Clin. Cancer Res.* **35**, 179 (2016).
- (BMBF #01KU1201A) and the German Cancer Aid (#109252) for the ICGC (International Cancer Genome Consortium) PedBrain Tumor Project and the ICGC MMML-Seq Project (within Program for Medical Genome Research Grants #01KU1002A–#01KU1002J), and the BioTOP Project (#01EK1502B). This work was also supported by an ERC starting grant to J.O.K. (#336045), MMML-MYC-SYS (#03161661), ICGC DE-Mining (#01KU1505G), the Heidelberg Center for Personalized Oncology (DKFZ-HIPO) and the BMBF-funded Heidelberg Center for Human Bioinformatics (HD-HuB) within the German Network for Bioinformatics Infrastructure (de.NBI) (#031A537A, #031A537C). For technical support and expertise we thank the DKFZ Genomics and Proteomics Core Facility, M. Hain from the Division of Molecular Genetics (DKFZ), N. Jaeger and R. Kabbe from the Department of Pediatric Neurooncology (DKFZ), and S. Oelmez from the Data Management Group (DKFZ). We further thank members and technical staff of the ICGC MMML-Seq (International Cancer Genome Consortium Molecular Mechanisms in Malignant Lymphoma by Sequencing) and the European Renal Tumor Study Group (SIOP-RTSG).
- Author Contributions** S.N.G. and B.C.W. performed data analysis and interpretation. S.N.G., J.W., I.B., K.K., V.A.R., G.P.B., M.S.-W., B.H., D.H., G.Z., M.H., J.E., C.L., and S.L. established workflows and performed data processing. P.D.J., S.Br., S.Be., D.S., E.P., S.E., S.W., U.K., J.J.M., G.V., C.P.K., M.Ko., D.T.W.J., L.C., and M.Z. contributed to design and interpretation of the analyses. P.D.J., D.H., C.B., A.B., M.Ku., S.F., J.W., R.K., D.B., A.E., S.Bu., R.K.-S., A.E.K., D.L., S.H., C.E., S.Bi., M.N., C.N., G.H.R., J.S., R.S., F.W., H.W., B.Bu., U.D., O.W., C.M.v.T., C.M.K., G.F., S.R., M.F., M.G., J.W., K.v.H., S.W., P.L., T.K., E.K., P.A.N., K.W.P., and M.Ko. provided data and patient materials. J.K., A.C.R., J.Z., Y.L., X.Z., A.J.W., D.A.Z., and P.R. established the databases. S.N.G., B.C.W., D.T.W.J. and S.M.P. prepared the manuscript and figures. B.Br., U.D.W., M.Ko., R.M.P., J.O.K., M.S., R.E., D.T.W.J., P.L., L.C., M.Z., and S.M.P. contributed to project management and provided leadership.
- Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to S.M.P. (s.pfister@dkfz.de).
- Reviewer Information** *Nature* thanks S. Chanock and the other anonymous reviewer(s) for their contribution to the peer review of this work.



This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) licence. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons licence, users will need to obtain permission from the licence holder to reproduce the material. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Susanne N. Gröbner^{1,2,3*}, Barbara C. Worst^{1,2,3,4*}, Joachim Weischenfeldt^{5,6}, Ivo Buchhalter⁷, Körtine Kleinheinz⁷, Vasilisa A. Rudneva^{5,8}, Pascal D. Johann^{1,2,3,4}, Gnana Prakash Balasubramanian^{1,2,9}, Maia Segura-Wang⁵, Sebastian Brabetz^{1,2,3}, Sebastian Bender^{1,2}, Barbara Hutter^{3,7,9}, Dominik Sturm^{1,2,3,4}, Elke Pfaff^{1,2,3,4}, Daniel Hübschmann^{4,9,10}, Gideon Zippich⁷, Michael Heindl^{7,10}, Jürgen Eils⁷, Christian Lawrenz⁷, Serap Erkek^{1,2,3,5}, Sander Lambo^{1,2,3}, Sebastian Waszak⁵, Claudia Blattmann^{3,11}, Arndt Borkhardt^{3,12}, Michaela Kühlen^{3,12}, Angelika Eggert^{3,13}, Simone Fulda^{3,14}, Manfred Gessler¹⁵, Jenny Wegert¹⁵, Roland Kappler^{3,16}, Daniel Baumhoer¹⁷, Stefan Burdach^{3,18}, Renate Kirschner-Schwabe^{3,13}, Udo Kontny^{3,19}, Andreas E. Kulozik^{1,3,4}, Dietmar Lohmann^{3,20}, Simone Hettner²¹, Cornelia Eckert^{3,13}, Stefan Bielack¹¹, Michaela Nathrath^{3,18,22}, Charlotte Niemeyer^{3,21}, Günther H. Richter^{3,18}, Johannes Schulte^{3,13}, Reiner Siebert²³, Frank Westermann^{3,24}, Jan J. Molenaar²⁵, Gilles Vassal²⁶, Hendrik Witt^{1,2,3,4}, ICGC PedBrain-Seq Project, ICGC MMML-Seq Project, Birgit Burkhardt²⁷, Christian P. Kratz²⁸, Olaf Witt^{1,3,4,29}, Cornelis M. van Tilburg^{1,3,30}, Christof M. Kramm³¹, Gudrun Fleischhack^{3,32}, Uta Dirksen³², Stefan Rutkowski³³, Michael Frühwald³⁴, Katja von Hoff³⁵, Stephan Wolf³⁵, Thomas Klingebiel^{3,36}, Ewa Koscielniak¹¹, Pablo Landgraf³⁷, Jan Koster³⁸, Adam C. Resnick³⁹, Jinghui Zhang⁴⁰, Yanling Liu⁴⁰, Xin Zhou⁴⁰, Angela J. Waanders⁴¹, Danny A. Zwijsburg³⁸, Pichai Raman³⁹, Benedikt Brors^{3,7,8}, Ursula D. Weber^{3,42}, Paul A. Northcott^{2,3,8}, Kristian W. Pajtl^{1,2,3,4}, Marcel Kool^{1,2,3}, Rosario M. Piro^{3,42,43,44}, Jan O. Korbel⁵, Matthias Schlesner^{7,45}, Roland Eils^{7,10}, David T. W. Jones^{1,2,3}, Peter Lichter^{3,42}, Lukas Chavez^{1,2,3,8}, Marc Zapatka^{42,43,8} & Stefan M. Pfister^{1,2,3,4} §

¹Hopp-Children's Cancer Center at the NCT Heidelberg (KITZ), Heidelberg, Germany. ²Division of Pediatric Neurooncology, German Cancer Research Center (DKFZ), Heidelberg, Germany.

³German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁴Department of Pediatric Oncology, Hematology & Immunology, Heidelberg University Hospital, Heidelberg, Germany. ⁵European Molecular Biology Laboratory (EMBL), Genome Biology Unit, Heidelberg, Germany. ⁶The Finsen Laboratory, Rigshospitalet, Biotech Research and Innovation Centre (BRIC), Copenhagen University, Copenhagen, Denmark.

⁷Division of Theoretical Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁸Department of Developmental Neurobiology, St Jude Children's Research Hospital,

Supplementary Information is available in the online version of the paper.

Acknowledgements This project was mainly supported and funded by the German Cancer Consortium (DKTK Pediatric Malignancies Joint Funding Project) and German Cancer Aid (#108128) and Deutsche Kinderkrebsstiftung (German Cancer Childhood Foundation) for the INFORM project. Additional support came from the German Ministry for Education and Research

Memphis, Tennessee, USA. ⁹Division of Applied Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg, Germany. ¹⁰Department of Bioinformatics and Functional Genomics, Institute of Pharmacy and Molecular Biotechnology, Heidelberg University and BioQuant Center, 69120, Heidelberg, Germany. ¹¹Klinikum Stuttgart - Olghospital, Zentrum für Kinder-, Jugend- und Frauenmedizin, Pädiatrie, Stuttgart, Germany. ¹²Department of Pediatric Oncology, Hematology & Clinical Immunology, University Children's Hospital, Heinrich Heine University, Düsseldorf, Germany. ¹³Department of Pediatric Oncology/Hematology, Charité-Universitätsmedizin Berlin, Berlin, Germany. ¹⁴Institute for Experimental Cancer Research in Pediatrics, University Hospital Frankfurt, Frankfurt am Main, Germany. ¹⁵Theodor-Boveri-Institute/BiCenter, Developmental Biochemistry, and Comprehensive Cancer Center Mainfranken, University of Würzburg, Würzburg, Germany. ¹⁶Department of Pediatric Surgery, Research Laboratories, Dr von Hauner Children's Hospital, Ludwig Maximilians University Munich, Munich, Germany. ¹⁷Bone Tumor Reference Center at the Institute of Pathology, University Hospital Basel and University of Basel, Basel, Switzerland. ¹⁸Children's Cancer Research Centre and Department of Pediatrics, Klinikum rechts der Isar, Technische Universität München, Munich, Germany. ¹⁹Division of Pediatric Hematology and Oncology, University Medical Center Aachen, Aachen, Germany. ²⁰Department of Human Genetics, University Hospital Essen, Essen, Germany. ²¹Division of Pediatric Hematology and Oncology, Department of Pediatrics, University Medical Center Freiburg, Freiburg, Germany. ²²Department of Pediatric Oncology, Klinikum Kassel, Kassel, Germany. ²³Institute of Human Genetics, University of Ulm & University Hospital of Ulm, Ulm, Germany. ²⁴Division of Neuroblastoma Genomics, German Cancer Research Center (DKFZ), Heidelberg, Germany. ²⁵Princess Máxima Center for Pediatric Oncology, Utrecht, The Netherlands. ²⁶Innovative Therapies for Children with Cancer Consortium and Department of Clinical Research, Gustave Roussy, Université Paris-Saclay, Villejuif, France. ²⁷Pediatric Hematology and Oncology, University Hospital Münster, Münster, Germany. ²⁸Pediatric Hematology and Oncology, Hannover Medical School, Hannover, Germany. ²⁹Clinical Cooperation Unit Pediatric Oncology, German Cancer Research Center (DKFZ), Heidelberg, Germany. ³⁰Center for Individualized Pediatric Oncology (ZIPO) and Brain Tumors, University Hospital and German Cancer Research Center (DKFZ), Heidelberg, Germany. ³¹Division of Pediatric Hematology and Oncology, University Medical Center Göttingen, Göttingen, Germany. ³²Pediatric Oncology & Hematology, Pediatrics III, University Hospital of Essen, Essen, Germany. ³³Department of Pediatric Hematology and Oncology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. ³⁴Swabian Children's Cancer Center, Children's Hospital, Klinikum Augsburg, Augsburg, Germany. ³⁵Genomics and Proteomics Core Facility, High Throughput Sequencing Unit, German Cancer Research Center (DKFZ), Heidelberg, Germany. ³⁶Hospital for Children and Adolescents, University Hospital Frankfurt, Frankfurt, Germany. ³⁷University Hospital Cologne, Klinik und Poliklinik für Kinder- und Jugendmedizin, Cologne, Germany. ³⁸Department of Oncogenomics, Academic Medical Center, Amsterdam, The Netherlands. ³⁹Division of Neurosurgery, Center for Childhood Cancer Research, Department of Biomedical and Health Informatics and Center for Data-Driven Discovery in Biomedicine, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. ⁴⁰Department of Computational Biology, St Jude Children's Research Hospital, Memphis, Tennessee, USA. ⁴¹Division of Oncology, Center for Childhood Cancer Research, Department of Biomedical and Health Informatics and Center for Data-Driven Discovery in Biomedicine, Children's Hospital of Philadelphia, Philadelphia, USA. ⁴²Division of Molecular Genetics, German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁴³Institute of Computer Science, Freie Universität Berlin, Berlin, Germany. ⁴⁴Institute of Medical Genetics and Human Genetics, Charité University Hospital, Berlin, Germany. ⁴⁵Bioinformatics and Omics Data Analysis, German Cancer Research Center (DKFZ), Heidelberg, Germany.

*These authors contributed equally to this work.

†These authors jointly supervised this work.

International Cancer Genome Consortium PedBrain Project

Coordination (WP1) Peter Lichter¹, Ursula Weber¹, Roland Eils²; **Banking (WP2)** Andrey Korshunov³, Olaf Witt^{4,5}, Stefan Pfister^{4,6}; **Reference Pathology & Quality Control (WP3)** Guido Reifenberger⁷, Jörg Felsberg⁷; **Isolation of Analytes (WP4)** Christof von Kalle⁸, Manfred Schmidt⁹, Cynthia Bartholomä⁹, Michael Taylor¹⁰; **Genomic Sequencing (WP5)** Stefan Pfister^{4,6}, David Jones^{4,6}, Peter Lichter¹, Natalie Jäger^{4,6}, Ivo Buchhalter²; **Paired-End Mapping (WP6)** Jan Korbel¹¹, Adrian Stütz¹¹, Tobias Rausch¹¹; **Methylome Sequencing (WP7)** Bernhard Radlwimmer¹; **Transcriptome Analysis (WP8)** Marie-Laure Yaspo¹², Hans Lehrach¹², Hans-Jörg Warnatz¹²; **Profiling of small non-coding RNAome (WP9)** Pablo Landgraf¹³, Arndt Borkhardt¹³; **Bioinformatics (WP10)** Benedikt Brors¹⁴, Marc Zapatka¹, Roland Eils²; **Bioinformatics (WP11)** Roland Eils², Jürgen Eils¹⁵ & Christian Lawrenz¹⁵

¹Division of Molecular Genetics, German Cancer Research Center (DKFZ), Heidelberg, Germany.

²Division of Theoretical Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg, Germany.

³Department of Neuropathology, Heidelberg University Hospital, Heidelberg, Germany.

⁴Hopp-Children's Cancer Center at the NCT Heidelberg (KITZ), Heidelberg, Germany.

⁵Clinical Cooperation Unit Pediatric Oncology, German Cancer Research Center (DKFZ), Heidelberg, Germany.

⁶Division of Pediatric Neurooncology, German Cancer Research Center (DKFZ), Heidelberg, Germany.

⁷Department of Neuropathology, Heinrich-Heine-University, Düsseldorf, Germany.

⁸Division of Translational Oncology, German Cancer Research Center (DKFZ)/National Center for Tumor Diseases (NCT), Heidelberg, Germany.

⁹GeneWerk GmbH, Heidelberg, Germany.

¹⁰Division of Neurosurgery, Hospital for Sick Children, Toronto, Ontario,

Canada. ¹¹Genome Biology Unit, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany. ¹²Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics (MPI-MG), Berlin, Germany. ¹³Hematology and Clinical Immunology, University Hospital, Düsseldorf, Germany. ¹⁴Division of Applied Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg, Germany. ¹⁵Data Management Group, German Cancer Research Center (DKFZ), Heidelberg, Germany.

International Cancer Genome Consortium MML-Seq Project

Coordination (C1) Reiner Siebert^{1,2}, Susanne Wagner², Andrea Haake², Julia Richter^{2,3}, Gesine Richter²; **Data Center (C2)** Roland Eils^{4,5}, Chris Lawrenz⁴, Jürgen Eils⁴, Jules Kerssemakers⁴, Christina Jaeger-Schmidt⁴, Ingrid Scholz⁴; **Clinical Centers (WP1)** Anke K. Bergmann^{2,6}, Christoph Borst⁷, Birgit Burkhardt^{8,9}, Alexander Claviez⁶, Martin Dreyling¹⁰, Sonja Eberth¹¹, Hermann Einsele¹², Norbert Frickhofen¹³, Siegfried Haas⁷, Martin-Leo Hansmann¹⁴, Dennis Karsch¹⁵, Michael Kneba¹⁵, Jasmin Lisfeld⁹, Luisa Mantovani-Löffler¹⁶, Marius Rohde⁹, German Ott¹⁷, Christina Stadler¹¹, Peter Staib¹⁸, Stephan Stilgenbauer¹⁹, Lorenz Trümper¹¹, Thorsten Zenz²⁰; **Normal Cells (WPN)** Martin-Leo Hansmann¹⁴, Dieter Kube¹¹, Ralf Küppers²¹, Marc Weniger²¹; **Pathology & Analyte Preparation (WP2-3)** Michael Hummel²², Wolfram Klapper³, Ulrike Kostezka²³, Dido Lenze²², Peter Möller²⁴, Andreas Rosenwald²⁵, German Ott¹⁷, Monika Szczepanowski³; **Sequencing & Genomics (WP4-7)** Ole Ammerpohl^{1,2}, Sietse M. Aukema^{2,3}, Vera Binder²⁶, Arndt Borkhardt²⁶, Andrea Haake², Jessica I. Hoell²⁶, Ellen Leich²⁵, Peter Lichter²⁷, Cristina López^{1,2}, Inga Nagel², Jordan Pischmarov²⁵, Bernhard Radlwimmer²⁷, Julia Richter^{2,3}, Philip Rosenstiel²⁸, Andreas Rosenwald²⁵, Markus Schilhabel²⁸, Stefan Schreiber²⁹, Inga Vater², Rabea Wagener^{1,2}, Reiner Siebert^{1,2}; **Bioinformatics (WP8-9)** Stephan H. Bernhart^{30,31,32}, Hans Binder^{30,31}, Benedikt Brors³³, Gero Doose^{30,31,32}, Roland Eils^{4,5}, Steve Hoffmann^{30,31,32}, Lydia Hopp³⁰, Daniel Hübschmann^{4,5,34}, Kortine Kleinheinz^{4,5}, Helene Kretzmer^{30,31,32}, Markus Kreuz³⁵, Jan Korbel³⁶, David Langenberger^{30,31,32}, Markus Loeffler³⁵, Maciej Rosolowski³⁵, Matthias Schlesner^{4,37}, Peter F. Stadler^{30,31,32,38,39,40} & Stephanie Sungalee³⁶

¹Institute of Human Genetics, University of Ulm and University Hospital of Ulm, Ulm, Germany.

²Institute of Human Genetics, Christian-Albrechts-University, Kiel, Germany. ³Hematopathology

Section, Institute of Pathology, Christian-Albrechts-University, Kiel, Germany. ⁴Division of

Theoretical Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg, Germany.

⁵Department for Bioinformatics and Functional Genomics, Institute of Pharmacy and Molecular

Biotechnology and Bioquant, University of Heidelberg, Heidelberg, Germany. ⁶Department of

Pediatrics, University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany. ⁷Department

of Internal Medicine/Hematology, Friedrich-Ebert-Hospital, Neumünster, Germany. ⁸University

Hospital Muenster - Pediatric Hematology and Oncology, Münster, Germany. ⁹University

Hospital Giessen, Pediatric Hematology and Oncology, Giessen, Germany. ¹⁰Department

of Medicine III - Campus Grosshadern, University Hospital Munich, Munich, Germany.

¹¹Department of Hematology and Oncology, Georg-August-University of Göttingen, Göttingen,

Germany. ¹²University Hospital Würzburg, Department of Medicine and Poliklinik II, University

of Würzburg, Würzburg. ¹³Department of Medicine III, Hematology and Oncology, Dr Horst-

Schmidt-Kliniken of Wiesbaden, Wiesbaden, Germany. ¹⁴Senckenberg Institute of Pathology,

University of Frankfurt Medical School, Frankfurt am Main, Germany. ¹⁵Department of Internal

Medicine II: Hematology and Oncology, University Medical Centre, Campus Kiel, Kiel, Germany.

¹⁶Hospital of Internal Medicine II, Hematology and Oncology, St-Georg Hospital Leipzig,

Leipzig, Germany. ¹⁷Department of Pathology, Robert-Bosch-Hospital, Stuttgart, Germany.

¹⁸Clinic for Hematology and Oncology, St-Antonius-Hospital, Eschweiler, Germany.

¹⁹Department for Internal Medicine III, University of Ulm and University Hospital of Ulm,

Ulm, Germany. ²⁰National Centre for Tumor Disease, Heidelberg, Germany. ²¹Institute of

Cell Biology (Cancer Research), University of Duisburg-Essen, Duisburg-Essen, Medical

School, Essen, Germany. ²²Institute of Pathology, Charité - University Medicine Berlin, Berlin,

Germany. ²³Comprehensive Cancer Center Ulm (CCCU), University Hospital Ulm, Ulm,

Germany. ²⁴Institute of Pathology, University of Ulm and University Hospital of Ulm, Ulm,

Germany. ²⁵Institute of Pathology, University of Würzburg, Würzburg, Germany. ²⁶Department

of Pediatric Oncology, Hematology and Clinical Immunology, Heinrich-Heine-University,

Düsseldorf, Germany. ²⁷German Cancer Research Center (DKFZ), Division of Molecular

Genetics, Heidelberg, Germany. ²⁸Institute of Clinical Molecular Biology, Christian-Albrechts-

University, Kiel, Germany. ²⁹Department of General Internal Medicine, University Kiel, Kiel,

Germany. ³⁰Interdisciplinary Center for Bioinformatics, University of Leipzig, Leipzig, Germany.

³¹Bioinformatics Group, Department of Computer, University of Leipzig, Leipzig, Germany.

³²Transcriptome Bioinformatics, LIFE Research Center for Civilization Diseases, University

of Leipzig, Leipzig, Germany. ³³Division of Applied Bioinformatics (G200), German Cancer

Research Center (DKFZ), Heidelberg, Germany. ³⁴Department of Pediatric Immunology,

Hematology and Oncology, University Hospital, Heidelberg, Germany. ³⁵Institute for Medical

Informatics Statistics and Epidemiology, University of Leipzig, Leipzig, Germany. ³⁶EMBL

Heidelberg, Genome Biology, Heidelberg, Germany. ³⁷Bioinformatics and Omics Data Analytics

(B240), German Cancer Research Center (DKFZ), Heidelberg, Germany. ³⁸RNomics Group,

Fraunhofer Institute for Cell Therapy and Immunology IZI, Leipzig, Germany. ³⁹Santa Fe

Institute, Santa Fe, New Mexico, USA. ⁴⁰Max-Planck-Institute for Mathematics in Sciences,

Leipzig, Germany.

METHODS

Samples. The cohort analysed in this study is a compilation of individual sequencing datasets from various sources: the International Cancer Genome Consortium (ICGC) – Pedbrain Tumor and MML-seq (<http://www.icgc.org>), the German Cancer Consortium (DKTK) (<https://dktk.dkfz.de/en/home>), the Pediatric Cancer Genome Project (PCGP) (<http://explore.pediatriccancergenomeproject.org/>), the Heidelberg Institute for Personalized Oncology (HIPO) (<http://www.dkfz.de/en/hipo>), the Individualized Therapy For Relapsed Malignancies in Childhood (INFORM) registry (www.dkfz.de/en/inform), and other previously published datasets (listed below). For all included tumours, matched germline control tissue was available. Ninety-five per cent of the patients were under 18 years of age (or age unspecified but confirmed age group paediatric), but available data were included for patients up to 25 years, as these were considered relevant for cancer types that typically peak at a young age. All centres have approved data access and informed consent had been obtained from all patients.

External data were downloaded from the European Genome-Phenome Archive (EGA; <https://www.ebi.ac.uk/ega/home>) using the accession numbers EGAD00001000085, EGAD00001000135, EGAD00001000159, EGAD00001000160, EGAD00001000161, EGAD00001000162, EGAD00001000163, EGAD00001000164, EGAD00001000165, EGAD00001000259, EGAD00001000260, EGAD00001000261, EGAD00001000268, and EGAD00001000269^{49–62}; internal datasets are related to previous PMIDs 27748748, 27479119, 26923874, 25670083, 25253770, 24972766, 24553142, 25135868, 26632267, 26179511, 24651015, 28726821, 23817572, 25962120, 26294725^{17,19,44,63–74} (Supplementary Note 1).

The final cohort included 914 individual patients of no more than 25 years of age including primary tumours for 879 patients with 47 matched relapsed tumours, and an additional 35 independent relapsed tumours (Supplementary Tables 1, 2). Deep-sequencing ($\sim 30\times$) whole-genome data (WGS) were available for 547 samples with matched control, whole-exome sequencing (WES) for 414, and low-coverage whole-genome sequencing (lcWGS) for an additional 54 germline and 186 tumour samples. Depending on the requirements of each sub-analysis, we used WES and WGS, WGS only (excluding Ewing's sarcoma, Wilms tumour, hepatoblastoma, and T-ALL), or WES, WGS and lcWGS (germline excluding Ewing's sarcoma, Wilms tumour and hepatoblastoma; tumours excluding Ewing's sarcoma and hepatoblastoma) were used (Supplementary Table 24). 'Subgroups' of cancer types were considered as separate entities if there was considerable evidence of differences in terms of clinical and molecular behaviours, if sub-cohort sizes were substantial, and if full annotation of all samples was available. All samples had been sequenced using Illumina technology and 99% of samples were paired-end sequences with 100 bp read length. Ninety-eight per cent of exome sequences are covered with at least $30\times$, 94% with at least $60\times$, and the total median exome coverage is $121\times$. The whole-genome sequenced samples have a median coverage of $37\times$ and 94% of samples are covered with at least $30\times$. Information on coverage and other metrics for all samples are provided in Supplementary Table 2.

Cancer type incidence. Information on incidence of cancer types in the population was derived from the SEER database (Surveillance, Epidemiology, and End Results program)⁸; further detailed information on different subgroups of cancer types (central nervous system tumours and subgroups of medulloblastoma, ependymoma, and ALL) was transferred from cancer type-specific publications^{75–79}. Survival data are based on information from the German Childhood Cancer Registry⁸⁰. Incidence rates of adult cancers were taken from information in the German GEKID database (<http://www.gekid.de/>, 2003–2012).

Data preprocessing. All data were processed using a standardized alignment and variant calling pipeline, which was developed in the context of the ICGC Pan-Cancer project (<https://dockstore.org/containers/quay.io/pancancer/pcawg-dkfz-workflow>)⁸¹.

Alignments. Datasets were available in either raw FASTQ or aligned BAM format. To allow standardized processing for all included samples, BAM files were sorted by read name using sambamba (v.0.4.6) and converted to a raw-like FASTQ format using SamToFastq (v.1.61). Reads were then aligned to the phase II reference human genome assembly of the 1000 Genomes Project including decoy sequences (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/phase2_reference_assembly_sequence/hs37d5.fa.gz) using BWA-MEM (v.0.7.8 using default settings except '-T 0'). Matching genotypes of tumour and control samples were confirmed by calculating pairwise DNA sequence similarities at 1,000 reference SNPs (dbSNP v.138)⁸².

Mutation calling. SNVs were called with the previously described samtools-based DKFZ pipeline adjusted for ICGC Pan-Cancer settings, and short indels were called using Platypus (v.0.7.4)^{74,83}. Variants were first identified in the tumour sample and germline or somatic origin was determined based on their presence or absence in the matched control tissue. Functional effects were annotated using ANNOVAR and GENCODE19 (<http://www.encodegenes.org/releases/19.html>)⁸⁴.

Somatic structural variant discovery. Somatic structural variant discovery was pursued across all whole-genome sequenced samples (high-quality structural variants available for $n = 539$ primary tumours) using the DELLY ICGC Pan-Cancer analysis workflow (https://github.com/ICGC-TCGA-PanCancer/pcawg_delly_workflow)⁸⁵. A high-stringency structural variant set was obtained by additionally filtering somatic structural variants detected in 1% or more of a set of 1,105 germline samples from healthy individuals belonging to phase I of the 1000 Genomes Project and by removing somatic structural variants present in any of the paediatric germline samples of this study⁸⁶. High-stringency structural variants were further required to have at least four supporting read pairs with a minimum mapping quality of 20 and were restricted to somatic structural variant sizes from 300 bp to 500 Mb.

Copy-number calling. Copy numbers were estimated using ACESeq (allele-specific copy-number estimation from sequencing) (K. Kleinheinz *et al.*, unpublished data), using a binned tumour–control coverage ratio and B-allele frequency (BAF). Allele frequencies were obtained for all single nucleotide polymorphism (SNP) positions recorded in dbSNP version 135⁸². To improve sensitivity with regard to imbalanced and balanced regions, SNP positions in the control were phased with impute2⁸⁷. Additionally, the coverage for 10-kb windows with sufficient mapping quality and read density was recorded and subsequently corrected for GC content and replication timing.

The genome was segmented using the PSCBS package incorporating structural variant breakpoints defined by DELLY^{88,89}. Segments were clustered based on coverage ratio and BAF using k -means and neighbouring segments in the same cluster were joined; focal segments (< 9 Mb) were stitched to the more similar neighbour. Tumour cell content and ploidy were estimated by testing how well different combinations of both explain the data. Segments with balanced BAF were assigned to even-numbered copy-number states, whereas unbalanced segments were allowed to match with uneven numbers as well. Finally, estimated tumour cell content and ploidy were used to compute the total and allele-specific copy-number for each segment. High-quality copy-number calls were available for $n = 516$ of the WGS samples.

Mutation statistics. The frequency of somatic mutations in coding regions was determined for each sample individually by normalizing the total number of coding mutations for the number of sufficiently covered ($\geq 6\times$) coding bases to account (determined using MuSiC-bmr) for different data types (WGS/WES) and for different exome target enrichment kits²⁴. Mutation spectra were obtained by categorizing observed SNVs into base substitution types in pyrimidine context. Spearman's rank correlation test was applied to infer correlations between different types of mutation counts or between mutation counts and age. Generalized linear models were used to fit regression lines. Clusters of localized hypermutation were identified using a previously presented approach adjusted for mutation rates in human paediatric cancers⁹⁰.

Deciphering mutation signatures. Exome-sequenced tumours, except for hypermutator cases, were excluded from signature analysis owing to their low numbers of mutations. In brief, signatures are represented as probability distributions of substitution types of SNVs in pyrimidine context. Considering the immediate sequence context of each SNV, this results in 96 possible mutation types with directly adjacent mutations (multiple nucleotide variants, MNVs) being excluded, which are counted per tumour to compile its mutational profile.

As proposed by Alexandrov *et al.*⁹¹, the mutational profile of a tumour is expected to reflect a superposition of mutational processes (signatures) acting on its genome, where each mutational process has a different intensity (exposure). For a cohort of tumour genomes, this is modelled as a system of matrices for signatures (P) and exposures (E) defining the observed mutational catalogue (M)⁹¹: $M \approx P \times E$.

De novo deciphering of signatures was done as described⁹¹ based on the mutational catalogues of all cancer types and of the pan-cancer cohort. All resulting signatures were compared to published signatures (available in the COSMIC database, <http://cancer.sanger.ac.uk/cosmic/signatures>) based on their cosine similarity¹⁵. Signatures that did not correspond to any of the previously known signatures (cosine similarity < 0.85) were further analysed to examine their relevance for modelling the cancer genomes. First, linear independence from the known set of signatures was confirmed. Second, for each potentially novel signature, we examined whether the modelling of mutation profiles improved when compared to having used the set of known signatures: for each sample, the observed mutational profile was compared to the theoretical profiles calculated using the set of known signatures only, and using the extended set including the new candidate signature. Here, only samples with a total number of mutations over 200 were considered. Reconstruction was calculated as the difference between cosine similarity of the modelled profile and the observed profile. On the basis of the resulting distribution of similarities in both alternatives, a signature was considered to have a relevant contribution to the model, and thus a potential new signature,

if both of the following conditions were fulfilled: the reconstruction (measured as the difference of similarities) of at least one sample increased by 0.02 and that sample had a reconstruction accuracy of <0.9 based on the known set of signatures only.

This procedure resulted in one new candidate signature, signature P1, which was added to the set of reference signatures. In order to achieve maximum resolution per sample, a sample-wise re-extraction of exposures from the mutational profiles was performed using quadratic programming with the reference signature set used for P and the exposures in E as unknown variables. Samples with a reconstruction accuracy below 0.5 were excluded (resulting in $n = 503$ tumours with high-quality signature information), as these samples would not be correctly accounted for by the model, which might be due to quality issues or to contributions of unknown signatures that are not present at intensities sufficient to be identified by a *de novo* approach. The resulting exposures were used for further downstream analyses and visualization. Previously published signatures without validation were first included to model the mutational catalogues as precisely as possible, but then summarized as 'other' for representation.

Spearman's rank correlation and two-sided Kolmogorov–Smirnov tests were used to associate exposure of signatures with numerical and categorical variables, respectively. Exposures to signatures across multiple groups were compared using ANOVA and the post hoc Tukey's test.

Identifying mutations in genes predisposing to cancers. To identify germline variants with a high likelihood of being implicated in cancer development, we investigated 162 candidate genes adapted from ref. 19 (110 genes regarded as following a dominant inheritance pattern and 52 genes with recessive inheritance) (Supplementary Table 6).

Germline SNVs and indels were subjected to a stepwise filtering approach to eventually classify them into five categories: benign, likely benign, uncertain significance, likely pathogenic, and pathogenic. First, variants reported in both the 1000 Genomes (release November 2010) and dbSNP (v.141) databases were excluded. High-quality variant calls were selected by including only positions with $\geq 15\times$ coverage, a germline allele frequency of ≥ 0.2 , and a phred-based quality score of ≥ 10 . Variants with a population frequency ≥ 0.01 reported in additional common databases (esp6500siv2, X1000g2015, and exac03 included in ANNOVAR (<http://annovar.openbioinformatics.org/>) or with ClinVar (<ftp://ftp.ncbi.nlm.nih.gov/pub/clinvar/>) annotations of 'benign', 'likely benign' or 'uncertain significance' were removed.

Furthermore, variants with a phred-scaled CADD score ≥ 15 (<http://cadd.gs.washington.edu/info>) and with Mutation Assessor (<http://mutationassessor.org/r3/>) categories 'medium' and 'high', or no available annotation, were included. Variants with a dbSNP classification of 'precious' were not subject to these two filtering steps. As indel calling is more prone to alignment and calling errors, potentially deleterious indels were manually investigated for artefacts. For recessive tumour genes, variants were included only with an allele frequency of one or with two compound heterozygous mutations of the same gene in the same patient. In total, the filtering steps narrowed down the number of potentially pathogenic mutations to $n = 433$. Every variant was then manually checked and scored by the use of varied, mainly gene-specific online databases (<http://p53.iarc.fr/>, <http://www.lovd.nl/3.0/home>, <https://www.ncbi.nlm.nih.gov/clinvar/>, and others). Only likely pathogenic and pathogenic mutations were considered as cancer-relevant and used for representation in Fig. 3. Additionally, whole-genome sequenced samples were manually screened for copy-number losses in 13 tumour suppressor genes of the candidate list, which are known to occasionally harbour germline focal deletions (*MLH1*, *MSH2*, *MSH6*, *NF1*, *PMS2*, *PRKAR1A*, *PTCH1*, *PTEN*, *RBI*, *SMARCA4*, *SMARCB1*, *SUFU*, *TP53*).

Detecting genome-wide mutation clusters. To identify genomic regions with single or clusters of recurrent mutations, the human genome was binned into non-overlapping windows of various sizes (50–500 bp) and compared the observed mutations to a background model (V. A. Rudneva *et al.*, unpublished data) which was estimated using the 'global' model: the genome was stratified into 25 evenly sized groups of genomic windows based on the combined vector of five genetic and epigenetic features (replication timing, gene expression level, GC content, H3K9me3, and open versus closed chromatin conformation). For each region an enrichment score, binomial P value, and negative binomial test P value were computed.

Cross-validations were used to determine the significance cut-off that would provide reproducible results (with samples segregated by subgroup). A combination of the window size (500 bp), test statistics (enrichment score, mutational recurrence, binomial test P value, and gamma Poisson test P value), and a cut-off value that ensured high precision and recall values based on the precision-recall analysis ($P = 10^{-20}$) were chosen (Extended Data Fig. 4a). Recall was calculated as the number of regions that satisfied the cut-off in results obtained on both halves of the dataset; precision was calculated as a fraction of the recalled regions to the total number of regions that satisfied the cut-off in each of the datasets. The chosen parameters were then used to run the pipeline on the complete dataset

and then the mutations in the resulting regions were further examined manually for potential false positives in order to identify high-confidence candidate regions (Extended Data Fig. 4b).

Significantly mutated genes. Significantly mutated genes based on somatic SNVs and indels were identified with the SMG module of the MuSiC tools suite²⁴ separately from all cancer types and from the pan-cancer cohort, and then merged.

This kind of significance analysis often produces false positive hits (for example, very large genes), despite normalization procedures, and thus several filters were applied to the raw output³⁰. First, all genes of $>30,000$ bp exonic length or $>10,000$ bp with additional replication timing >800 were excluded (Cancer Cell Line Encyclopedia; CCLE)⁹². Genes that scored significant in three or more cancer types, or that were recurrently mutated at the same position, were manually inspected for artefacts from ambiguous alignments (for example, repetitive sequence regions). Also, genes that are probably not associated with tumour development but rather represent non-neoplastic somatic hypermutation processes in the context of immune function were removed. Furthermore, genes mutated in $<2\%$ of the cohort were included only if they had a secondary signal from either functional impact or from localized clustering bias (Intogen modules OncodriveFM and OncodriveClust v. 3.0 beta) or from being among known cancer genes^{29,93}. Mutation needle plots were generated using MutationMapper⁹⁴. Biological processes were assigned to the significantly mutated genes mostly exclusively, except for a few genes with high relevance for multiple processes, as specified in Supplementary Table 9.

Genome instability. Occurrence of chromothripsis was determined by manual inspection of coverage ratio plots (tumour/control) for WGS samples based on previously proposed guidelines⁹⁵: at least ten copy-number switches on one chromosome, oscillating copy-number variation (usually with changes of $+1$ or -1 , but also between other levels where additional large-scale copy-number changes interfere), and many more of such copy-number variations in one chromosome or chromosome arm compared to the remaining genome. In samples with an exceptionally high degree of structural variation, several chromosomes could be affected, and some samples showed an 'amplifier' type of chromothripsis, which was classified as several high-level focal amplifications on exactly the same copy-number level that are thus likely to be connected to one single event.

Generation of copy-number profiles. Copy-number calls reported by ACESeq were converted to the 'SEG' segmentation format, similar to the output of the circular binary segmentation algorithm based on chromosomal segment borders as pseudo marker positions⁹⁶. All possible marker positions were determined from the whole cohort before assessing sample-wise copy-number profiles per marker in order to achieve identical resolution for all samples. Owing to sparse and highly oscillating sequencing coverage at centromeres, centromeric coordinates (± 3 Mb around the centre of annotated centromeres) were excluded from whole-genome segmentation, as were two likely artefact regions on chromosomes 7 and 14 with nonspecific occurrences of relative copy-number gains and losses in 28% and 30% of all analysed samples in 17 of 19 entities (14q11.2, 7p14.1), which were identified using GISTIC2.0 (as described below) with ± 1 Mb.

Identifying recurrent copy-number/structural variations. GISTIC2.0 (v.2.0.22, gene-gistic default parameter settings) was applied to the segmented copy-number data (per cancer type and pan-cancer) to identify significant copy-number alterations³⁶. The resulting peaks were filtered for significance ($q \leq 0.1$) and size (≤ 10 Mb). Compared to array-based data, which commonly serve as inputs for copy-number significance analysis, sequencing-based copy-number profiles are more prone to artefact copy-number variations, for example, due to repetitive regions leading to ambiguous alignments. Thus, several filtering steps were used to eliminate false-positive GISTIC peak calls and to discover potentially cancer-relevant copy-number alterations: first, peaks overlapping with common fragile genomic sites were excluded, as these are likely to be consequences of genomic instability rather than cancer-driving events⁹⁷; next, peaks overlapping within 1 Mb of chromosomal ends were removed, as here sequencing coverage tends to vary frequently; and last, peaks overlapping with copy-number variable regions⁹⁸ (regions ranked 1–100) were excluded. Additionally, some of the resulting peaks were classified as 'passengers' of variable regions that were called as separated peaks from most likely one event, for example, a peak with *MYCN*OS as passenger peak of *MYCN* amplification. For overlapping peaks called in multiple entities and/or pan-cancer, the final region was determined based on the analysis with highest significance for each peak, respectively.

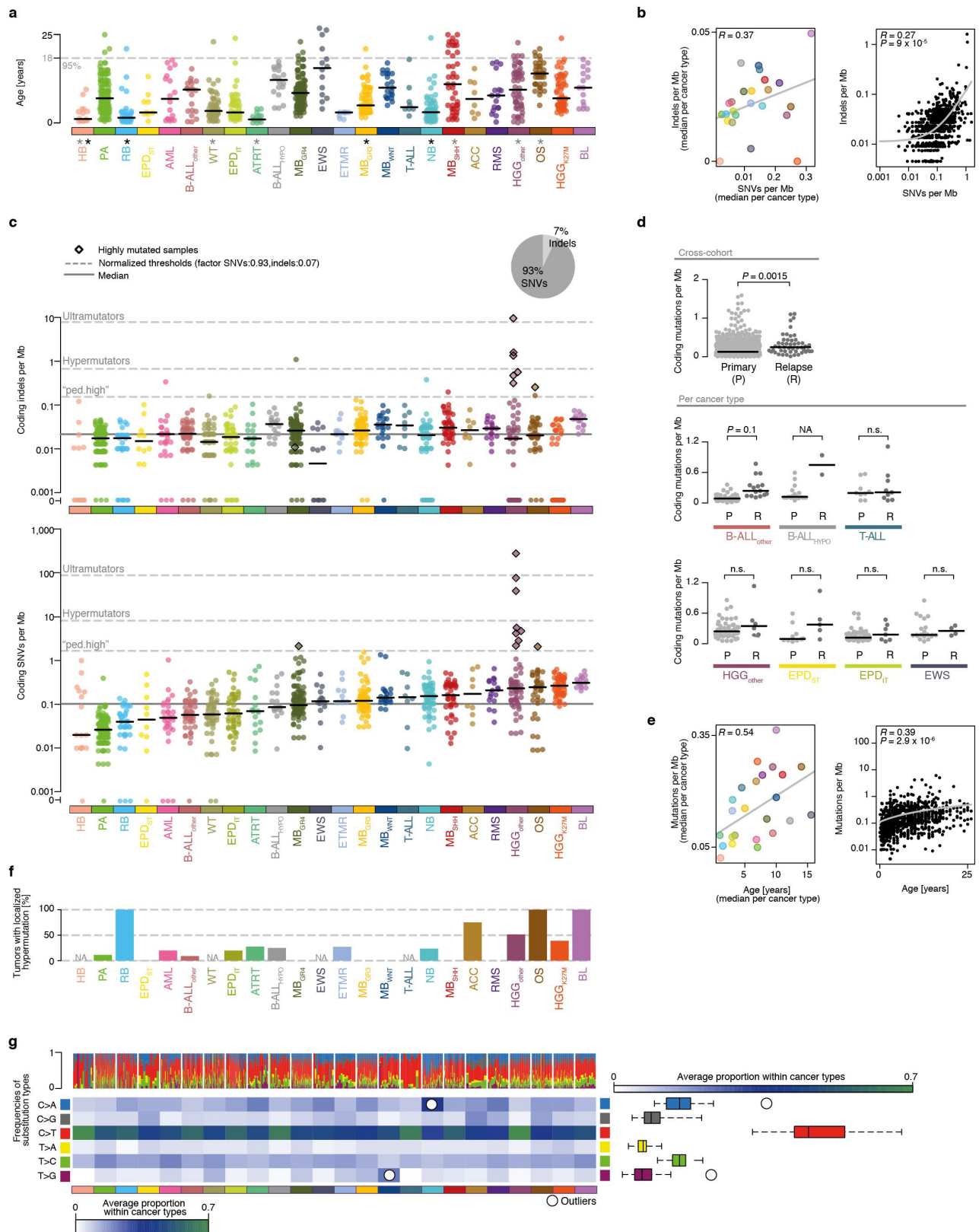
Genes with a breakpoint inside the gene borders were assumed to be altered by structural variation and considered as recurrently altered if they had breakpoints in ≥ 5 samples in total or in ≥ 2 samples of one cancer type (for samples without chromothripsis). For other samples, genes with breakpoints in ≥ 5 samples were included as candidates, but these were not used for further downstream analyses. Additionally, recurrent sites of structural variation outside of gene bodies by clustering breakpoints were determined in 10-kb windows.

Scoring of druggable mutations. To identify candidates for targeted therapy, somatic and germline mutations (SNV and indels) were screened for variants in genes that are directly or indirectly involved in pathways with matched drugs either approved or currently being investigated in clinical trials (Supplementary Table 22a, adapted from ref. 19). The mutations were then manually assessed by experts in translational oncology and prioritized according to an internal algorithm taking into account the type of alteration, the mechanism of action of potential drugs within the pathway, the level of evidence for the specific alteration, and its role in the present cancer type (Supplementary Table 22b, adapted from ref. 19). Only alterations scored 'intermediate' or 'high' were regarded as being relevant in terms of druggability. A clonality analysis was not performed owing to limited sequencing depth in whole-genome-sequenced tumours.

Additionally, copy-number plots of whole-genome-sequenced data (including low-coverage WGS) were used to manually screen 52 druggable genes for amplifications or deletions (Supplementary Table 22a). Only focal CNVs (<10 Mb) with at least 5 copies ($\log_2 \geq 1.3$) in the case of amplifications or the loss of ≥ 1 copy ($\log_2 \leq -1$) for deletions were included and subsequently prioritized as described for the SNVs/indels. The data representation includes all tumours with full genomic information (WES + lcWGS or WGS; $n = 675$) and, additionally, tumours analysed by WES only for cancer types without any whole-genome-sequenced tumours (T-ALL, Ewing's sarcoma, HB; $n = 39$), but the latter were excluded from downstream analyses.

Data availability. Mutation data have been deposited into commonly used public data portals and are accessible at <http://pedpancan.com>. They can be explored in and downloaded from the R2 Analysis and Genomics Platform, the PedcBio Portal for Cancer Visualization, and the TARGET Data Matrix. Sequencing data were obtained from previous studies as listed in Supplementary Note 1 and include the following accession codes: RP012816, PRJEB11430 (European Nucleotide Archive); EGAS00001001139, EGAS00001001953, EGAS00001000607, EGAS00001000381, EGAS00001000906, EGAS00001001297, EGAS00001000443, EGAS00001000213, EGAS00001000263, EGAS00001000192, EGAS00001000255, EGAS00001000254, EGAS00001000253, EGAS00001000256, EGAS00001000246, EGAS00001000379, EGAS00001000380, EGAS00001000346, EGAS00001000349, EGAS00001000347, EGAS00001000192 (European Genome-Phenome Archive).

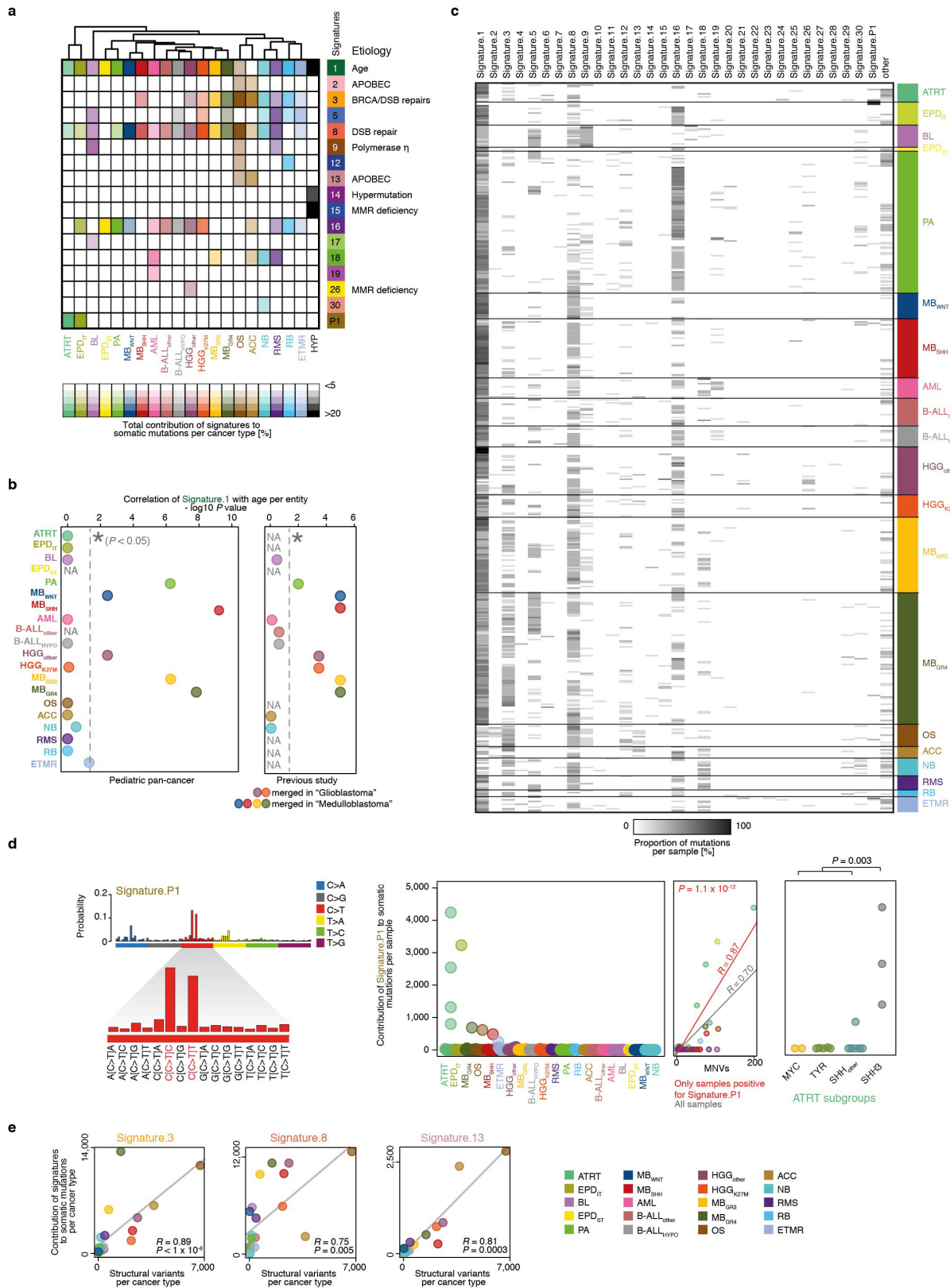
49. Wu, G. *et al.* Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. *Nat. Genet.* **44**, 251–253 (2012).
50. Wu, G. *et al.* The genomic landscape of diffuse intrinsic pontine glioma and pediatric non-brainstem high-grade glioma. *Nat. Genet.* **46**, 444–450 (2014).
51. Cheung, N. K. *et al.* Association of age at diagnosis and genetic mutations in patients with neuroblastoma. *J. Am. Med. Assoc.* **307**, 1062–1071 (2012).
52. Chen, X. *et al.* Recurrent somatic structural variations contribute to tumorigenesis in pediatric osteosarcoma. *Cell Rep.* **7**, 104–112 (2014).
53. Pinto, E. M. *et al.* Genomic landscape of paediatric adrenocortical tumours. *Nat. Commun.* **6**, 6302 (2015).
54. Zhang, J. *et al.* Whole-genome sequencing identifies genetic alterations in pediatric low-grade gliomas. *Nat. Genet.* **45**, 602–612 (2013).
55. Parker, M. *et al.* C11orf95-RELA fusions drive oncogenic NF- κ B signalling in ependymoma. *Nature* **506**, 451–455 (2014).
56. Chen, X. *et al.* Targeting oxidative stress in embryonal rhabdomyosarcoma. *Cancer Cell* **24**, 710–724 (2013).
57. Andersson, A. K. *et al.* The landscape of somatic mutations in infant MLL-rearranged acute lymphoblastic leukemias. *Nat. Genet.* **47**, 330–337 (2015).
58. Gruber, T. A. *et al.* An Inv(16)(p13.3q24.3)-encoded CBFA2T3-GLIS2 fusion protein defines an aggressive subtype of pediatric acute megakaryoblastic leukemia. *Cancer Cell* **22**, 683–697 (2012).
59. Holmfeldt, L. *et al.* The genomic landscape of hypodiploid acute lymphoblastic leukemia. *Nat. Genet.* **45**, 242–252 (2013).
60. Zhang, J. *et al.* A novel retinoblastoma therapy from genomic and epigenetic analyses. *Nature* **481**, 329–334 (2012).
61. Faber, Z. J. *et al.* The genomic landscape of core-binding factor acute myeloid leukemias. *Nat. Genet.* **48**, 1551–1556 (2016).
62. Robinson, G. *et al.* Novel mutations target distinct subgroups of medulloblastoma. *Nature* **488**, 43–48 (2012).
63. International Cancer Genome Consortium PedBrain Tumor Project. Recurrent MET fusion genes represent a drug target in pediatric glioblastoma. *Nat. Med.* **22**, 1314–1320 (2016).
64. Wegert, J. *et al.* Mutations in the SIX1/2 pathway and the DROSHA/DGCR8 miRNA microprocessor complex underlie high-risk blastemal type Wilms tumors. *Cancer Cell* **27**, 298–311 (2015).
65. Irving, J. *et al.* Ras pathway mutations are prevalent in relapsed childhood acute lymphoblastic leukemia and confer sensitivity to MEK inhibition. *Blood* **124**, 3420–3430 (2014).
66. Bandapalli, O. R. *et al.* The activating STAT5B N642H mutation is a common abnormality in pediatric T-cell acute lymphoblastic leukemia and confers a higher risk of relapse. *Haematologica* **99**, e188–e192 (2014).
67. Mack, S. C. *et al.* Epigenomic alterations define lethal CIMP-positive ependymomas of infancy. *Nature* **506**, 445–450 (2014).
68. Eichenmüller, M. *et al.* The genomic landscape of hepatoblastoma and their progenies with HCC-like features. *J. Hepatol.* **61**, 1312–1320 (2014).
69. Kovac, M. *et al.* Exome sequencing of osteosarcoma reveals mutation signatures reminiscent of BRCA deficiency. *Nat. Commun.* **6**, 8940 (2015).
70. Agelopoulos, K. *et al.* Deep sequencing in conjunction with expression and functional analyses reveals activation of FGFR1 in Ewing sarcoma. *Clin. Cancer Res.* **21**, 4935–4946 (2015).
71. Kool, M. *et al.* Genome sequencing of SHH medulloblastoma predicts genotype-related response to smoothened inhibition. *Cancer Cell* **25**, 393–405 (2014).
72. Li, B. *et al.* Negative feedback-defective PRPS1 mutants drive thiopurine resistance in relapsed childhood ALL. *Nat. Med.* **21**, 563–571 (2015).
73. Kunz, J. B. *et al.* Pediatric T-cell lymphoblastic leukemia evolves into relapse by clonal selection, acquisition of mutations and promoter hypomethylation. *Haematologica* **100**, 1442–1450 (2015).
74. Jones, D. T. *et al.* Recurrent somatic alterations of FGFR1 and NTRK2 in pilocytic astrocytoma. *Nat. Genet.* **45**, 927–932 (2013).
75. Ostrom, Q. T. *et al.* Alex's Lemonade Stand Foundation infant and childhood primary brain and central nervous system tumors diagnosed in the United States in 2007–2011. *Neuro-oncol.* **16** (Suppl 10), x1–x36 (2015).
76. Pajtler, K. W. *et al.* Molecular classification of ependymal tumors across all CNS compartments, histopathological grades, and age groups. *Cancer Cell* **27**, 728–743 (2015).
77. Northcott, P. A. *et al.* Medulloblastomics: the end of the beginning. *Nat. Rev. Cancer* **12**, 818–834 (2012).
78. Harrison, C. J. *et al.* Three distinct subgroups of hypodiploidy in acute lymphoblastic leukaemia. *Br. J. Haematol.* **125**, 552–559 (2004).
79. Pui, C. H., Relling, M. V. & Downing, J. R. Acute lymphoblastic leukemia. *N. Engl. J. Med.* **350**, 1535–1548 (2004).
80. Kaatsch, P. S. C. *German Childhood Cancer Registry - Report 2013/14 (1980–2013)* (Institute of Medical Biostatistics, Epidemiology and Informatics (IMBEI), Univ. Medical Center of Johannes Gutenberg Univ., 2014).
81. Stein, L. D., Knoppers, B. M., Campbell, P., Getz, G. & Korbel, J. O. Data analysis: Create a cloud commons. *Nature* **523**, 149–151 (2015).
82. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).
83. Jones, D. T. *et al.* Dissecting the genomic complexity underlying medulloblastoma. *Nature* **488**, 100–105 (2012).
84. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
85. Rausch, T. *et al.* DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**, i333–i339 (2012).
86. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
87. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).
88. Olshen, A. B. *et al.* Parent-specific copy number in paired tumor-normal studies using circular binary segmentation. *Bioinformatics* **27**, 2038–2046 (2011).
89. Van den Meersche, K., Soetaert, K. & Van Oevelen, D. xsample(): An R function for sampling linear inverse problems. *J. Stat. Softw.* **30**, 1–15 (2009).
90. Roberts, S. A. *et al.* Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. *Mol. Cell* **46**, 424–435 (2012).
91. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep.* **3**, 246–259 (2013).
92. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
93. Futreal, P. A. *et al.* A census of human cancer genes. *Nat. Rev. Cancer* **4**, 177–183 (2004).
94. Vohra, S. & Biggin, P. C. Mutationmapper: a tool to aid the mapping of protein mutation data. *PLoS ONE* **8**, e71711 (2013).
95. Korbel, J. O. & Campbell, P. J. Criteria for inference of chromothripsis in cancer genomes. *Cell* **152**, 1226–1236 (2013).
96. Olshen, A. B., Venkatraman, E. S., Lucito, R. & Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5**, 557–572 (2004).
97. Le Tallec, B. *et al.* Common fragile site profiling in epithelial and erythroid cells reveals that most recurrent cancer deletions lie in fragile sites hosting large genes. *Cell Rep.* **4**, 420–428 (2013).
98. Sudmant, P. H. *et al.* An integrated map of structural variation in 2,504 human genomes. *Nature* **526**, 75–81 (2015).



Extended Data Figure 1 | Somatic mutation frequencies and spectra.

a, Patient age at diagnosis (black lines, median); asterisks highlight cancer types with significant correlation of mutation load with age within cancer types (grey, SNVs; black, indels). **b**, Correlation of SNV and indel loads (left, median per cancer type; right, cross-cohort, $n = 876$). **c**, Somatic mutation frequencies (top, indels; bottom, SNVs) in the coding region ($n = 879$) (black lines, median). **d**, Mutation loads in primary versus relapse tumours (cross-cohort $n = 958$, per cancer type, see Supplementary Table 1; two-sided t -test, confidence interval 0.95). n.s., not significant;

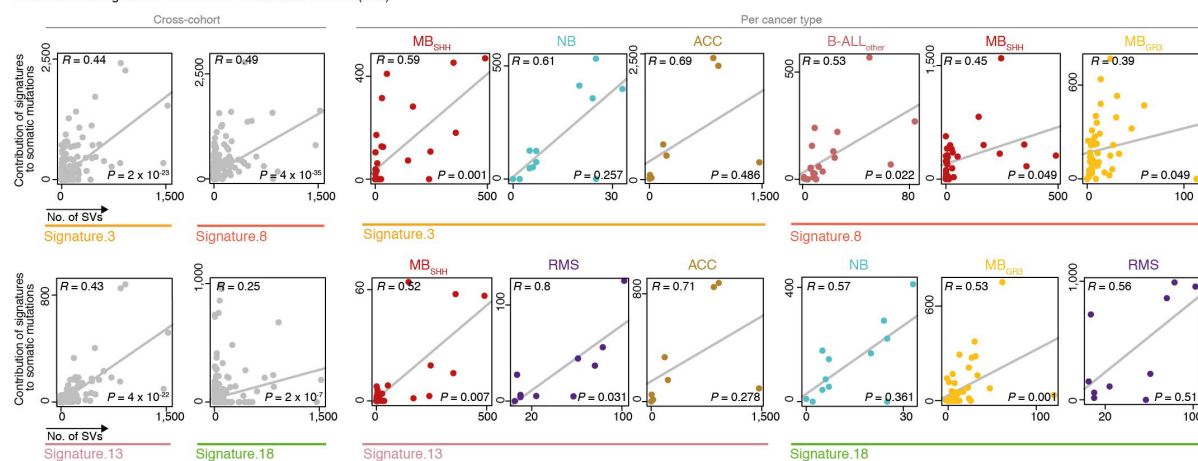
NA, not applicable. **e**, Correlation of mutations (SNVs and indels) with age (left, median per cancer type; right, cross-cohort $n = 876$). **f**, Proportion of tumours with one or several events of localized hypermutation (WGS samples, $n = 540$). **g**, Mutation spectra of SNVs (top, per sample; bottom, average per cancer type; $n = 879$). Distributions of frequencies per substitution type across cancer types are indicated on the right; outliers are highlighted in the heat map (quartiles, range of whiskers: $1.5 \times$ interquartile range). **a**, **b**, **e**, Linear model, confidence interval 0.95. Hypermutators and ultramutators are considered only in **c**.



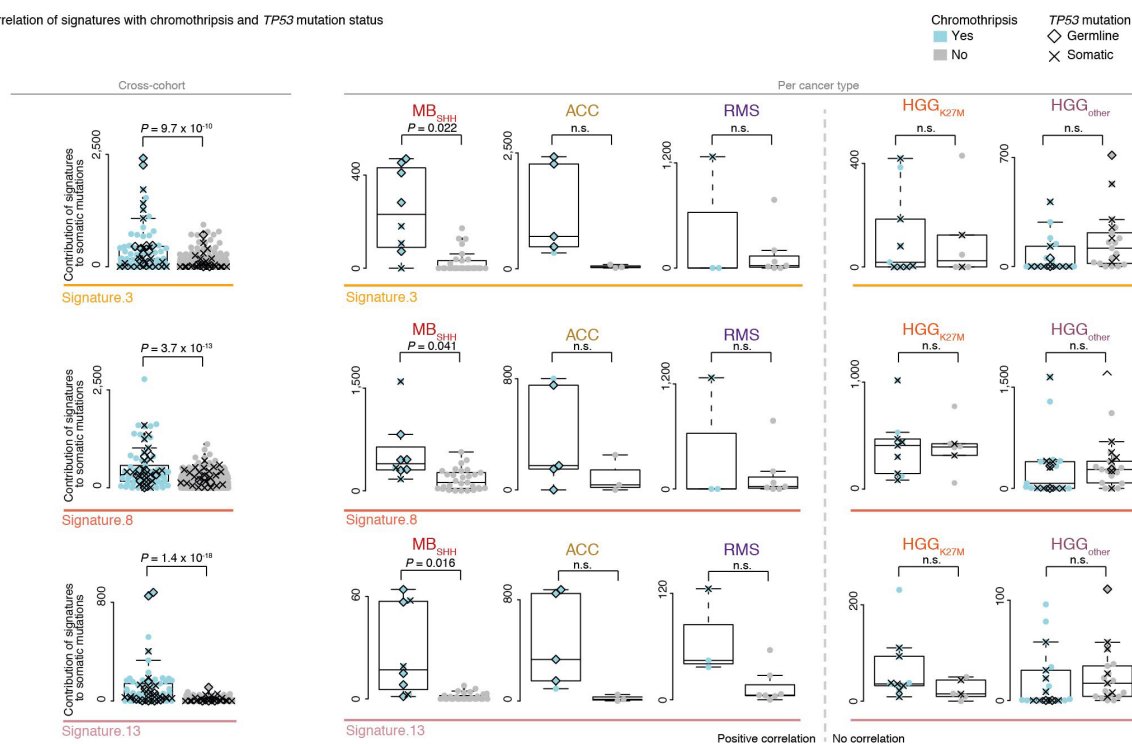
Extended Data Figure 2 | Mutational signatures in paediatric cancer types. **a**, Summarized contribution of signatures to mutational profiles per cancer type (proportion of mutations per signature and cancer type). Signatures with contributions of $\geq 5\%$ in at least one cancer type are shown. The colour intensity reflects the relative activity of each signature per cancer type. **b**, Correlation of signature 1 with patient age per cancer type in this paediatric pan-cancer cohort (left, $n = 503$) compared to results from a global pan-cancer study on 30 cancer types ($n = 7,042$)¹⁵. **c**, Relative contributions of mutational signatures to somatic mutations per

individual tumour, clustered within cancer types ($n = 503$). **d**, Correlation of signatures 3, 8, and 13 (somatic mutations) with genome instability (structural variants) per cancer type. **e**, Substitution type probabilities in trinucleotide context for the newly discovered mutational signature P1; contribution of signature P1 per tumour ($n = 503$); correlation of signature P1 with multiple nucleotide variants (MNVs); activity of signature P1 in ATRT subgroups (Wilcoxon rank-sum test, confidence interval 0.95). **b–d**, Spearman's correlation, confidence interval 0.95.

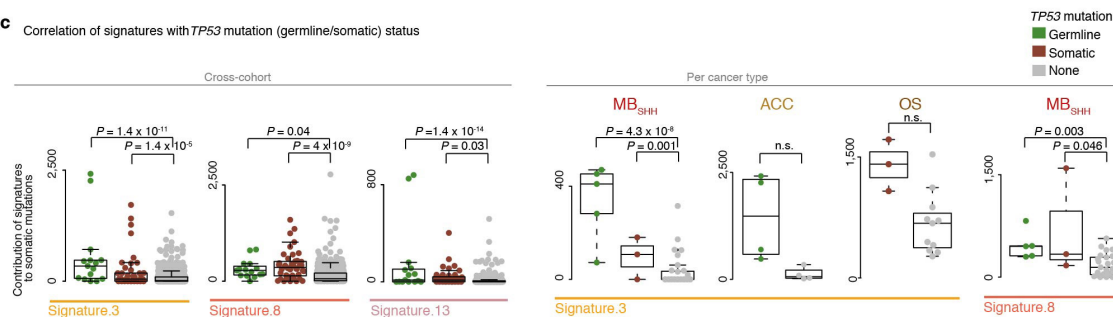
a Correlations of signatures with number of structural variants (SVs)



b Correlation of signatures with chromothripsis and *TP53* mutation status

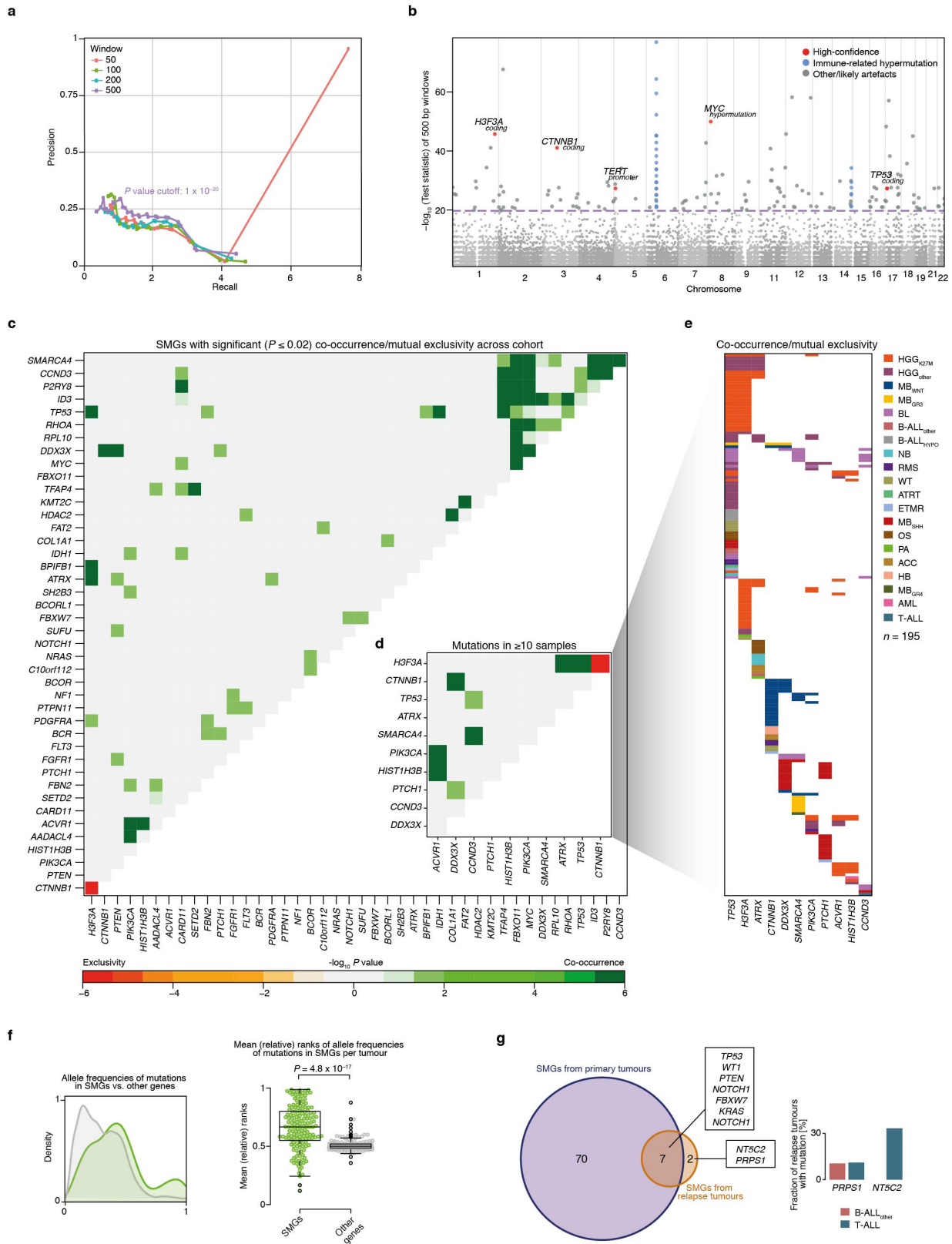


c Correlation of signatures with *TP53* mutation (germline/somatic) status



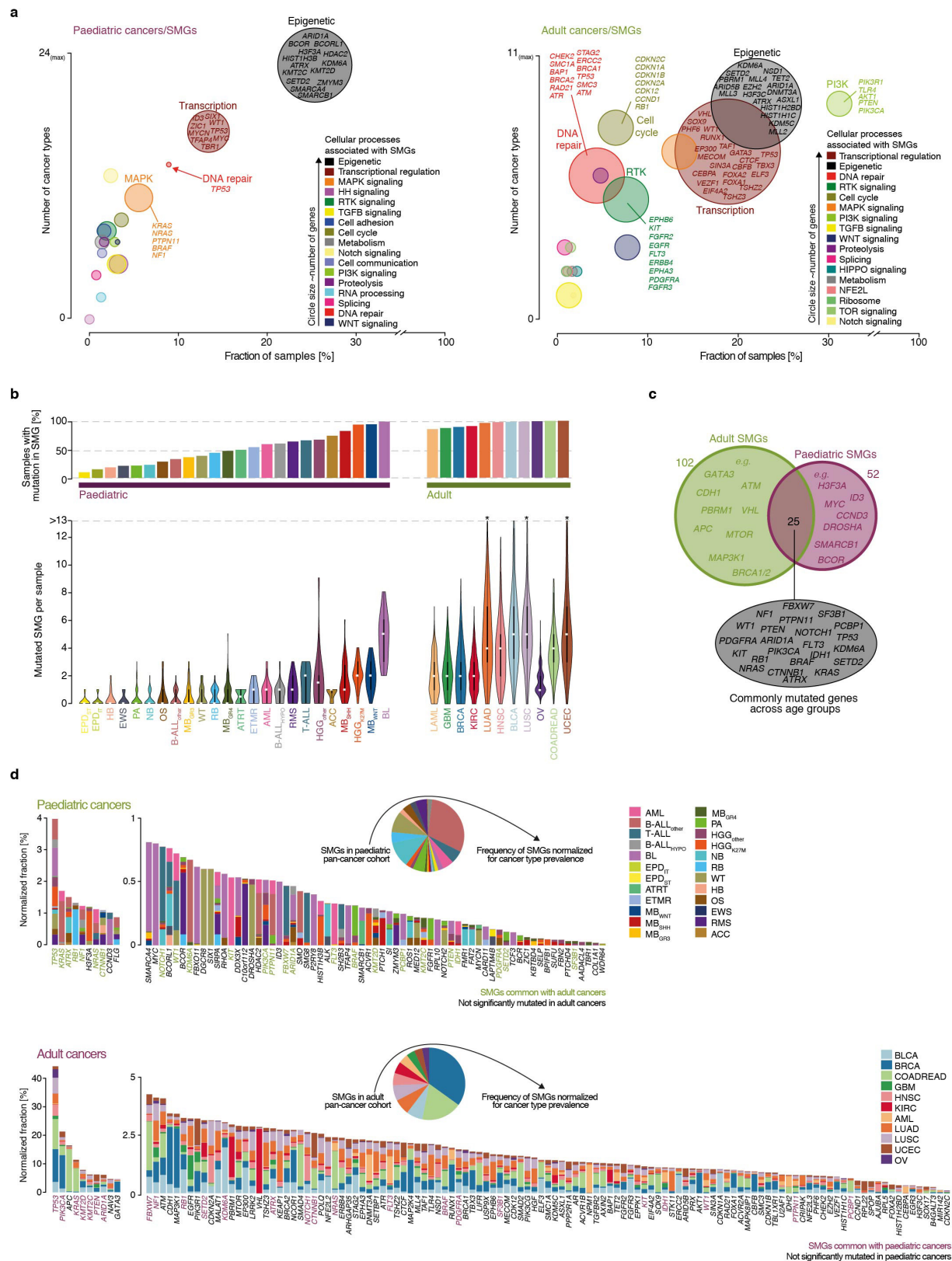
Extended Data Figure 3 | Association of mutational signatures with genomic instability. **a**, Correlation of signatures with the number of structural variants across all tumours and selected cancer types (Spearman's correlation, confidence interval 0.95). **b**, Association of signatures with chromothripsis across all tumours and within selected cancer types. *TP53* mutation status (germline/somatic) is highlighted

(Kolmogorov-Smirnov test, confidence interval 0.95, range of whiskers: $1.5 \times$ interquartile range). **c**, Association of signatures with *TP53* mutation status (germline/somatic/none) across all tumours and within selected cancer types (ANOVA and post hoc Tukey's test, confidence interval 0.95, quartiles, range of whiskers: $1.5 \times$ interquartile range). **a–c**, Cross-cohort $n = 503$, cancer types see Supplementary Tables 1, 4.



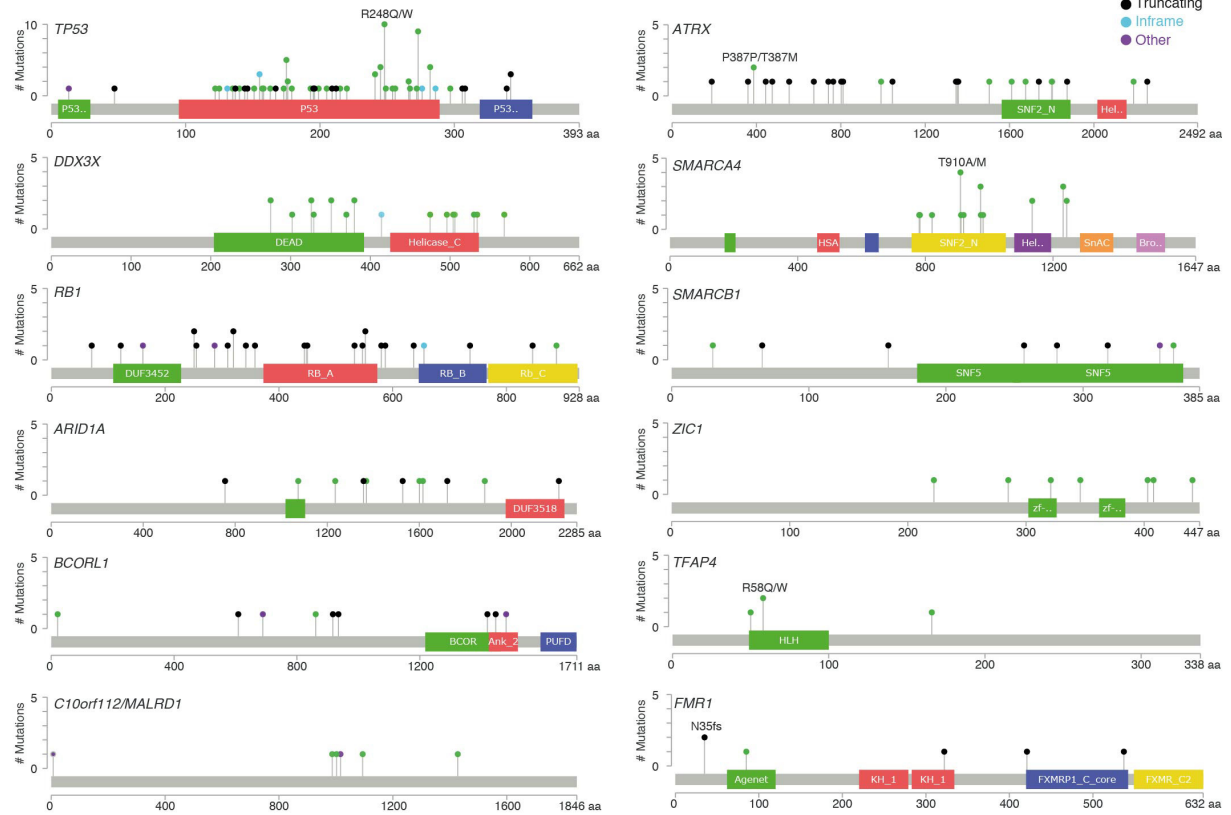
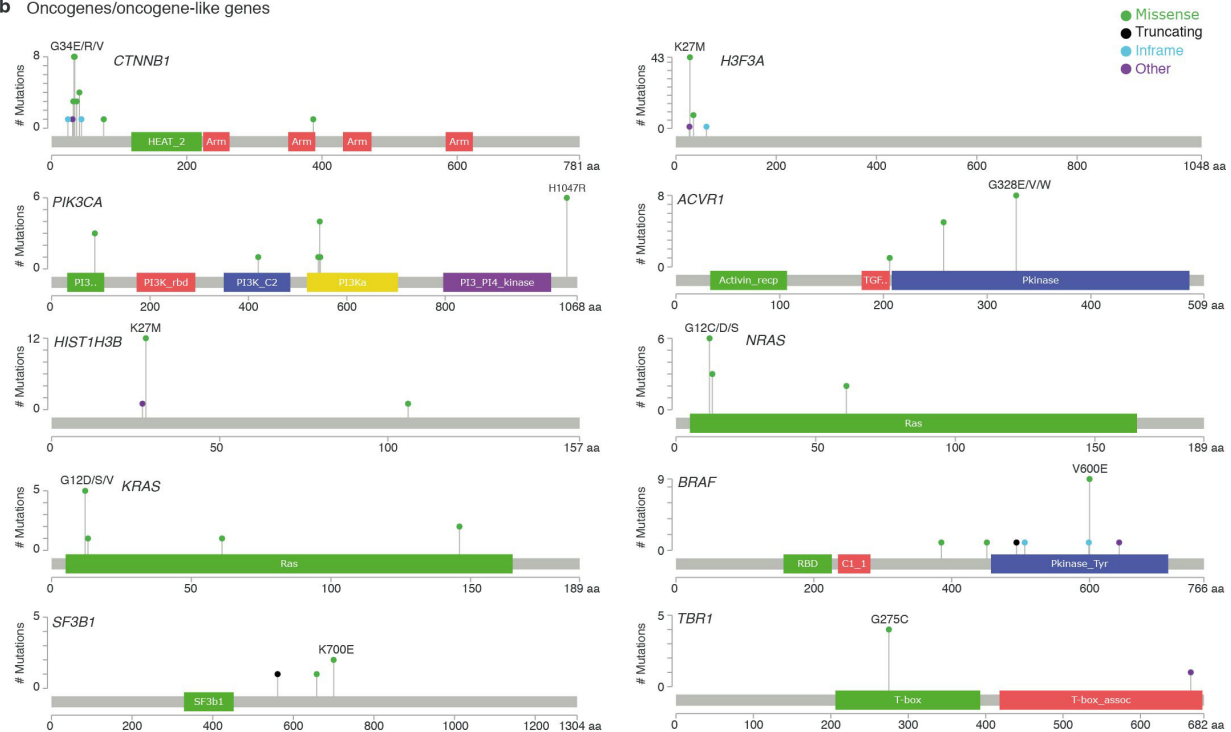
Extended Data Figure 4 | Characteristics of significantly mutated genomic regions and genes. **a**, Precision-recall curves (mean precision) for various binomial P value cut-offs for the identification of genome-wide mutation clusters. **b**, Manhattan plot for the test statistic of genomic windows. Dashed line indicates the P value cutoff from **a**. **c**, Significant co-occurrence/mutual exclusivity of SMGs in the pan-cancer dataset

($n = 876$). **d**, Most frequently mutated genes from **c**. **e**, Mutations in SMGs selected in **d** per cancer type. **f**, Allele frequencies of mutations in SMGs compared to mutations in non-SMGs in $n = 876$ tumours (two-sided t -test, confidence interval 0.95, quartiles, range of whiskers: $1.5 \times$ interquartile range). **g**, SMGs identified from relapse tumours and representation in cancer types.



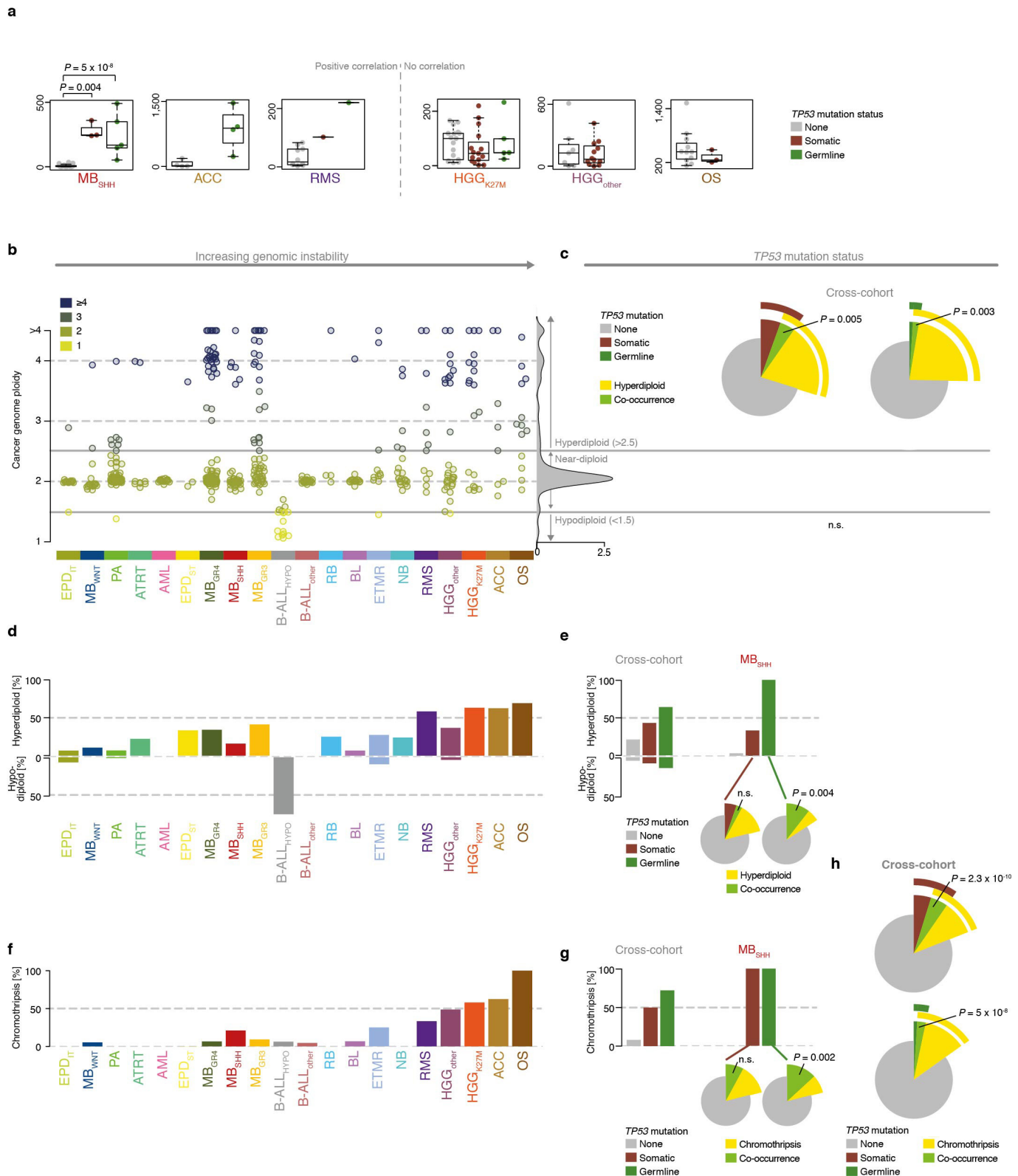
Extended Data Figure 5 | Significantly mutated genes across age groups. **a**, Cellular processes associated with paediatric (left) and adult (right) SMGs. **b**, Frequency of mutations in SMGs in paediatric ($n = 879$) compared to adult ($n = 3,281$) cancers. Top, percentage of SMG-mutated samples. Bottom, mutations in SMGs per sample (centre, median; range,

minimum to maximum). **c**, Overlap of SMGs detected in paediatric and adult cancers. **d**, Projected mutation rates of SMGs based on normalization of the cohort frequencies for cancer type incidence among patients for paediatric and adult cancers. **a–d**, Information on adult SMGs is based on TCGA data and previous analysis⁷.

a Tumor suppressor (-like) genes**b** Oncogenes/oncogene-like genes

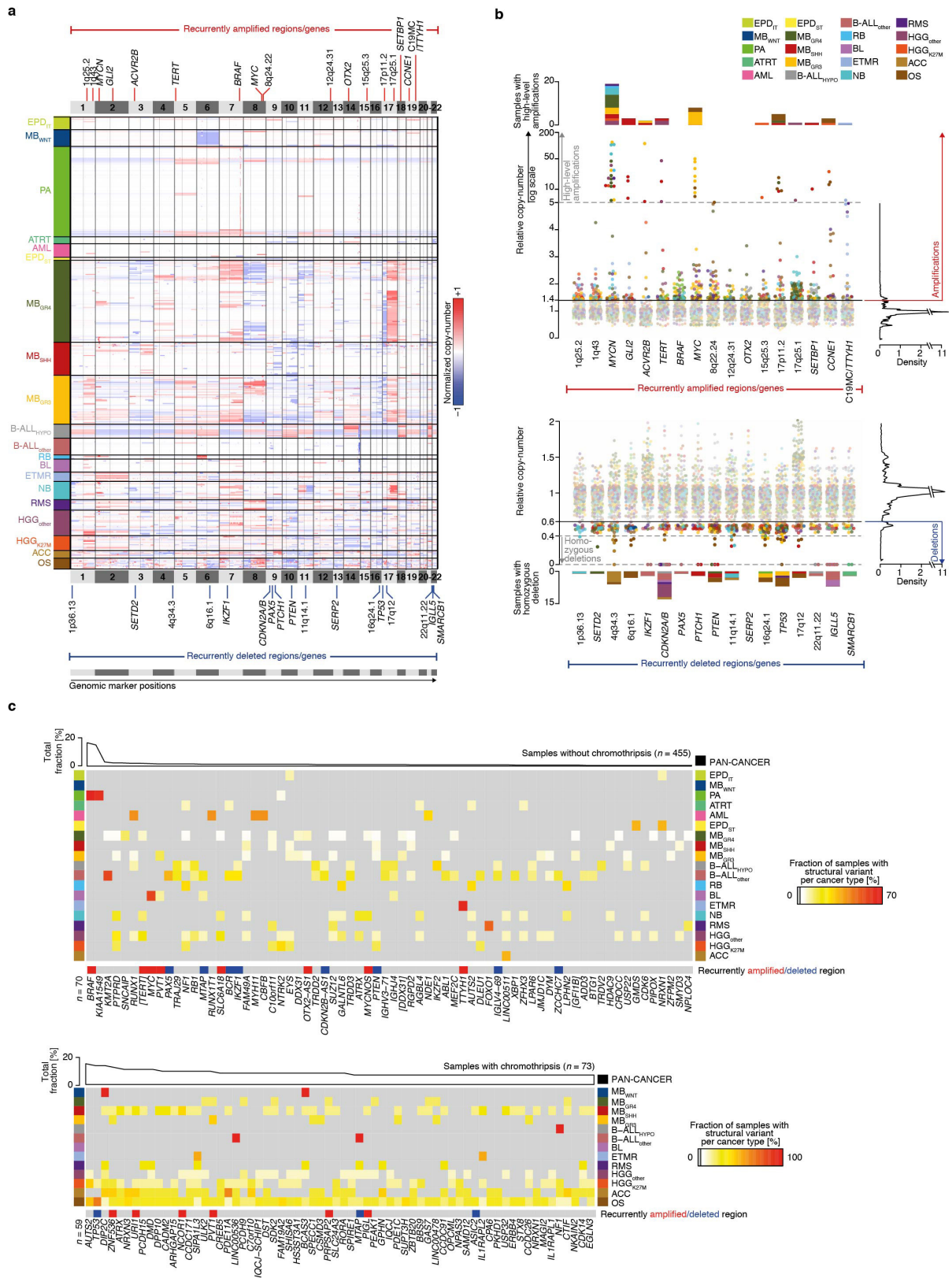
Extended Data Figure 6 | Mutation needle plots for significantly mutated genes. Mutations in selected significantly mutated genes across pan-cancer cohort: missense (green), truncating (black), in-frame (blue),

and other (purple). Hotspot amino acid changes are highlighted. **a**, Genes with tumour suppressor-like mutation patterns. **b**, Genes with oncogenic or oncogene-like mutation patterns.



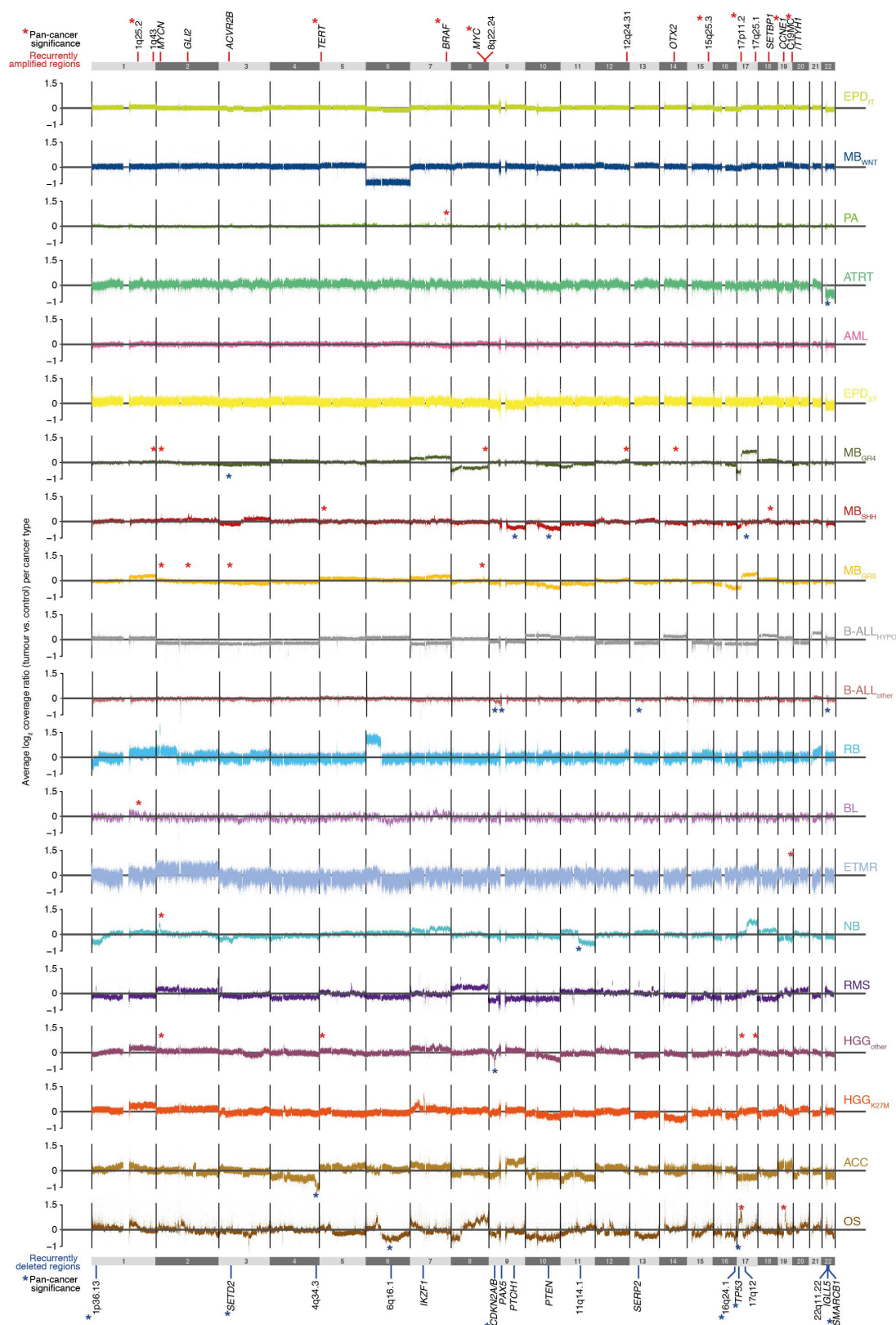
Extended Data Figure 7 | Genomic instability across paediatric cancer types. **a**, Structural variant load in relation to *TP53* mutation status for individual cancer types (generalized linear model, confidence interval 0.95). **b–h**, Characteristics of genomic instability (left) and their associations with *TP53* mutation status (right) (n.s., not significant). **b**, Genome ploidy; density of ploidy across all lineages is summarized on the right. **c**, Co-occurrence (Fisher's exact test) of hyperdiploidy (cross-cohort, $n = 516$) and *TP53* mutations (left, somatic; right, germline). **d**, Percentage of tumours per cancer type with hyper- (≥ 1.5) and

hypodiploid (≤ 0.5) genomes. **e**, Rate of hypodiploidy in relation to *TP53* mutation status (left, cross-cohort; right, cancer type-specific ($n_{SHH} = 38$) with co-occurrence highlighted as in **b**). **f**, Rate of chromothripsis (positive/negative). **g**, Rate of chromothripsis in relation to *TP53* mutation status (left, cross-cohort; right, cancer type-specific ($n_{SHH} = 38$) with co-occurrence highlighted as in **b**). **h**, Cross-cohort ($n = 516$) co-occurrence of samples with chromothripsis and *TP53* mutations (top, somatic; bottom, germline).



Extended Data Figure 8 | Recurrent CNVs and structural variations. **a**, Genome-wide copy-number profiles normalized for tumour ploidy ($n = 516$). Cancer types are sorted by genome instability (Fig. 5a). Regions or genes with significant CNVs are indicated (blue, deleted; red, gained or amplified) (Fig. 5b). **b**, Relative copy-number status (normalized for tumour ploidy to baseline 1) for regions with significant copy-number changes (top, gains or amplifications; bottom, deletions) in $n = 516$ tumours. Thresholds (amplified: ≥ 1.4 , deleted: ≤ 0.6) are based on the

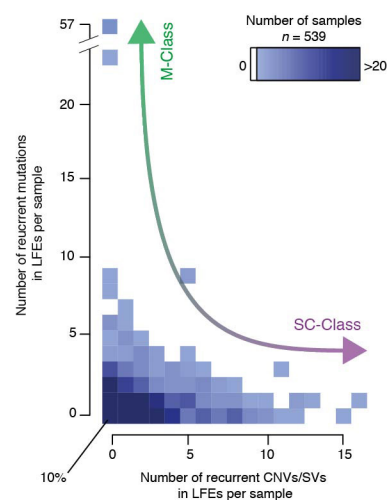
overall copy-number distribution indicated on the right. **c**, Genes affected by breakpoints from structural variants and additional genes associated with clustered breakpoints (in square brackets). Samples are divided into sub-cohorts of tumours with (bottom, $n = 73$) and without (top, $n = 455$) chromothripsis. Genes overlapping (direct overlap or within ± 200 kb) with genes with significant copy-number changes from **a** (blue, deletions; red, amplifications).



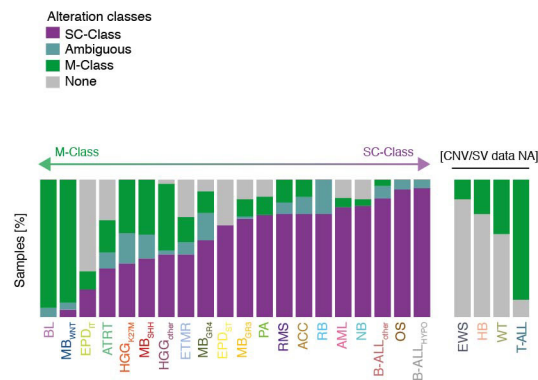
Extended Data Figure 9 | Averaged copy-number profiles per cancer type. Averaged copy-number profiles for all cancer types ordered by genome instability (Fig. 5a) and significant regions (Fig. 5b). The x-axis represents chromosomal positions in 1-kb windows and the y-axis the log₂

coverage of tumours versus controls. Asterisks indicate in which cancer types a region was called significant (amplifications, red and above copy-number profiles; deletions, blue and below profiles).

b



d



Tumours with more than 50% (mixed) or 100% (unique) events from one category are considered to be members of the associated class; tumours with equal contributions from both categories are 'ambiguous', and tumours without an LFE are assigned class 'none' (not shown). Colours indicate germline mutations per tumour. **d**, Fraction of tumours assigned to different classes per cancer type.

Tumours with more than 50% (mixed) or 100% (unique) events from one category are considered to be members of the associated class; tumours with equal contributions from both categories are 'ambiguous', and tumours without an LFE are assigned class 'none' (not shown). Colours indicate germline mutations per tumour. **d**, Fraction of tumours assigned to different classes per cancer type.

indicate germline mutations per tumour. **d**, Fraction of tumours assigned to different classes per cancer type.

The atomic structure of a eukaryotic oligosaccharyltransferase complex

Lin Bai¹, Tong Wang², Gongpu Zhao³, Amanda Kovach¹ & Huilin Li¹

N-glycosylation is a ubiquitous modification of eukaryotic secretory and membrane-bound proteins; about 90% of glycoproteins are N-glycosylated. The reaction is catalysed by an eight-protein oligosaccharyltransferase (OST) complex that is embedded in the endoplasmic reticulum membrane. Our understanding of eukaryotic protein N-glycosylation has been limited owing to the lack of high-resolution structures. Here we report a 3.5 Å resolution cryo-electron microscopy structure of the *Saccharomyces cerevisiae* OST complex, revealing the structures of subunits Ost1–Ost5, Stt3, Wbp1 and Swp1. We found that seven phospholipids mediate many of the inter-subunit interactions, and an Stt3 N-glycan mediates interactions with Wbp1 and Swp1 in the lumen. Ost3 was found to mediate the OST–Sec61 translocon interface, funnelling the acceptor peptide towards the OST catalytic site as the nascent peptide emerges from the translocon. The structure provides insights into co-translational protein N-glycosylation, and may facilitate the development of small-molecule inhibitors that target this process.

N-glycosylation is a predominant modification of proteins in eukaryotic organisms^{1–3}. N-linked glycans serve many essential functions, affecting protein folding and sorting in the endoplasmic reticulum (ER) and mediating interactions of the cell or organism with its environment⁴. Proteins are N-glycosylated in the ER lumen by the oligosaccharyltransferase complex, which transfers a pre-formed 14-sugar oligosaccharide from a dolichol-linked donor to selected asparagine residues within a conserved sequon, NXS/T (in which X can be any amino acid except proline)^{5,6}. Almost two-thirds of proteins include the NXS/T sequon and 65–75% of them are glycoproteins^{4,7}. Mutations in the OST proteins cause a family of diseases known as congenital disorders of glycosylation⁸.

The prokaryotic OST complex is a single-subunit enzyme that is homologous to the eukaryotic catalytic subunit Stt3⁹. The structures of the bacterial PglB and the archaeal AglB provided first insights into the glycosyl transfer reaction^{10–13}. However, prokaryotic protein N-glycosylation is simpler and occurs post-translationally. By contrast, most N-glycosylation in eukaryotes is co-translational¹⁴. Eukaryotes have evolved a sophisticated machinery to cope with this complexity. *Saccharomyces cerevisiae* has two OST isoforms each with eight membrane proteins: the isoforms contain either Ost3 or Ost6 plus seven shared components: Ost1, Ost2, Ost4 and Ost5; Stt3; Wbp1; and Swp1¹⁵. All of these subunits have homologues in the metazoan OST complex²: ribophorin I corresponds to yeast Ost1, DAD1 to Ost2, N33/MagT1 or DC2/KCP2 to Ost3/Ost6, OST4 to Ost4, TMEM258 to Ost5, OST48 to Wbp1, STT3A/STT3B to Stt3, and ribophorin II to Swp1¹⁶. Crystal structures of the Ost6 luminal domain revealed a thioredoxin (TRX) fold^{17,18}. The structures of Ost4 were solved by NMR^{19,20}. Biochemical studies suggested that Ost1 and Wbp1 recognize acceptor and donor substrates, respectively^{8,21,22}. The structures of the eukaryotic OST complex have been limited to low-resolution electron microscopy reconstructions, hindering a mechanistic understanding of protein N-glycosylation in eukaryotes^{23–26}.

Overall architecture of the OST

OST was purified from yeast strain LY510 (Methods). Purified OST is mainly of isoform Ost3, as Ost6 was barely detectable (Extended Data

Fig. 1). We determined a 3.5 Å resolution cryo-electron microscopy (cryo-EM) three-dimensional map and built an atomic model (Fig. 1a–c, Extended Data Figs 2, 3, Extended Data Table 1, Supplementary Videos 1, 2). The model contains 4 out of the 5 luminal domains, 26 out of the 28 transmembrane helices (TMHs), 3 N-glycans at Asn336 of Ost1p, Asn60 of Wbp1p, and Asn539 of Stt3p, and 8 phospholipids.

All five OST soluble domains are in the ER lumen (Fig. 1b, c). The four well-resolved soluble domains are from Stt3, Ost1, Wbp1 and Swp1. The fifth domain is from Ost3, which has TRX activity and interacts with the nascent peptide¹⁷. This domain is visible in one 3D class, indicating flexibility probably owing to the absence of the acceptor peptide (Extended Data Fig. 4). These domains are arranged in an intermediate layer proximal to the membrane and a top layer distal to the membrane. In the intermediate layer, the luminal domain of Stt3 binds tightly with the middle domain of Wbp1 and NTD2 of Ost1. By contrast, the three domains in the top layer, NTD of Wbp1, NTD1 of Ost1, and NTD of Swp1, are packed loosely.

Notably, the transmembrane domain of the OST complex has a triangular shape in which Stt3 is in the centre, surrounded by all other subunits (Extended Data Fig. 5a–c). TMH2–TMH4 of Ost3 pack against TMH10–TMH11 and TMH13 of Stt3, forming the top angle. At the lower right angle, TMH1–TMH3 of Ost2 directly interacts with TMH5 and TMH7–TMH8 of Stt3, and TMH1 of Wbp1 and TMH1–TMH3 of Swp1 are further out, interacting with Ost2. TMH1 of Ost1 and TMH1–TMH2 of Ost5 are loosely organized and constitute the lower left angle. The two missing TMHs, TMH9 of Stt3 and TMH1 of Ost3, surround the lipid-linked oligosaccharide (LLO) docking site and are presumably flexible in the absence of the donor.

The structure of Stt3

The catalytic subunit Stt3 comprises 13 TMHs, a luminal domain, and an α -helical accessory domain that is formed by external loop 1 (EL1) (Fig. 2a, b). The 13-TMH topology is consistent with a published characterization of yeast and mouse Stt3 and is similar to the prokaryotic homologues^{13,27}. The Stt3 luminal domain is a mixed α/β -fold, similar to the prokaryotic counterparts^{10–13,28}, but it is different from the NMR structure of the yeast Stt3 that was derived from proteins refolded in

¹Center for Epigenetics, Van Andel Research Institute, Grand Rapids, Michigan, USA. ²Advanced Science Research Center at the Graduate Center of the City University of New York, New York, New York, USA. ³David Van Andel Advanced Cryo-Electron Microscopy Suite, Van Andel Research Institute, Grand Rapids, Michigan, USA.

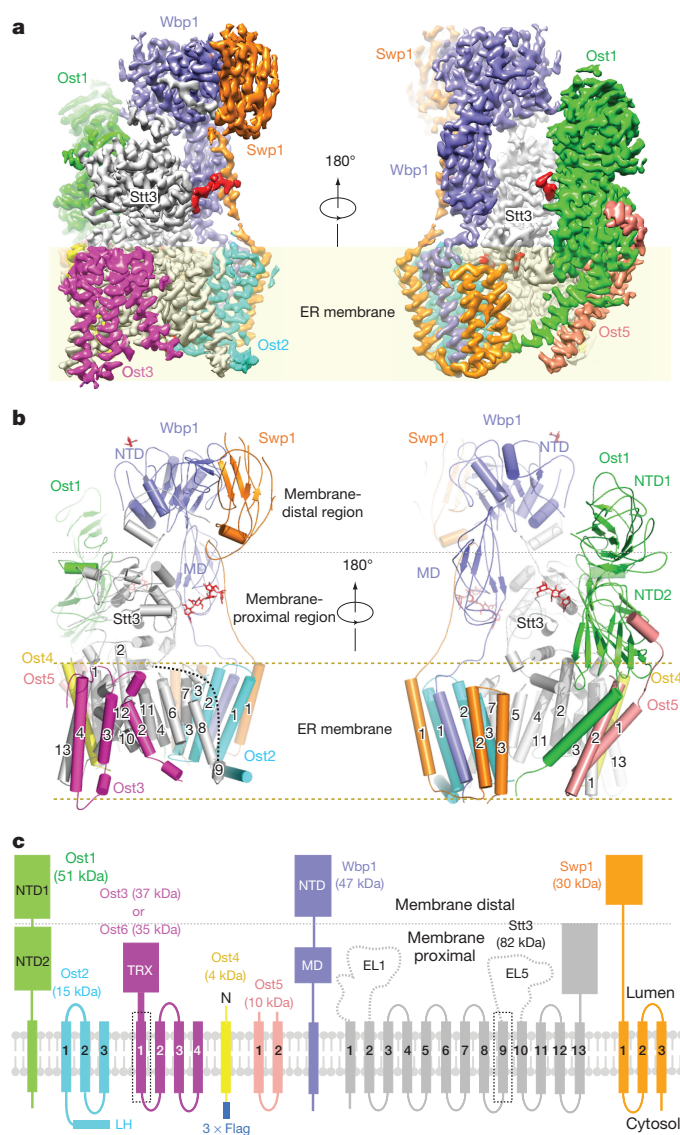


Figure 1 | Subunit composition and atomic structure of the yeast OST complex. **a**, Cryo-EM 3D map is shown in front and back views and coloured by individual subunits. The shaded pale yellow rectangle represents the ER membrane. The three N-glycan densities are in red. **b**, Atomic structure is shown in cartoons. Three N-linked glycans are displayed as sticks. The grey dotted line separates the membrane proximal and membrane-distal luminal regions. **c**, The domain structures of the eight subunits. 'N' denotes the N terminus of Ost4. EL1 and EL5 mark the external loops 1 and 5 in the Stt3 transmembrane domain. Two flexible TMHs are highlighted by dotted squares.

the presence of the denaturing sodium dodecyl sulfate²⁹. The structural conservation between Stt3 and prokaryotic enzymes is remarkable, considering their approximately 20% sequence identity. The bacterial enzymes have an extended acceptor sequon of DXNXS/T³⁰. The –2 position D is stabilized by R331 in the PglB crystal structure¹³. PglB R331 is replaced by D362 in yeast Stt3, explaining the shorter eukaryotic sequon of NXN/T. Notably, the yeast Stt3 has a carboxyl terminal extension (CTE) that mediates the interaction between Wbp1 and Swp1 (Fig. 2a, c, Extended Data Fig. 3). The Stt3 CTE is apparently a eukaryote-specific feature as it is absent in prokaryotes (Extended Data Figs 6, 7). The CTE of the metazoan STT3A is shorter than that of the STT3B. It would be interesting to investigate whether the longer CTE endows the STT3B isoform with the capacity to glycosylate post-translationally the extreme C-terminal acceptor sites of the nascent polypeptides that have been skipped by the STT3A isoform³¹.

There is a horizontal substrate-binding groove in Stt3 between the transmembrane domain and luminal domain. The roof of the groove is lined by the highly conserved glycosylation-sequon-binding WWD motif. Compared with the periplasmic domain of PglB in complex with a peptide substrate, the Stt3 luminal domain is shifted outwards by 3 Å, leading to a widened groove (Fig. 2b), which may facilitate substrate binding. Relative to PglB, TMH1 and TMH13 in Stt3 move away from each other to accommodate the sole TMH of Ost4; in this regard, Ost4 may be considered as an integral part of Stt3 (Extended Data Fig. 5c). TMH8–TMH9 slide by 25 Å towards the LLO-binding surface, leaving a space for Ost2 TMH1–TMH2. Stt3 residues that are expected to coordinate Mn²⁺ and LLO, such as Asp47, Asp166, Glu168, Trp208 and Arg404, are superimposable with their respective counterparts in PglB (Fig. 2b, Extended Data Fig. 6b).

The Stt3 N-glycan in the Asn539–Asn540–Asn541 sequon interacts with Asp320, Tyr322 and Arg349 of Wbp1 and with His182 in the linker loop of Swp1 (Fig. 2a, d), suggesting that the glycan stabilizes the complex by gluing these subunits together. A previous study showed that the Asn539Gln mutation is lethal, and the Thr541Ala mutation is severely temperature-sensitive³².

Ordered phospholipids mediate OST complex assembly

The eight OST subunits can be grouped into three subcomplexes: Ost1–Ost5, Ost2–Swp1–Wbp1 and Stt3–Ost3–Ost4 (Fig. 3a, b). This architecture is consistent with a previous cross-linking study³³. The interfaces among the three subcomplexes in the membrane region are loose, with only a few protein–protein interactions between Ost2 and Stt3. We identified seven phospholipids (PL1–PL7) at the interfaces of these subcomplexes, and an eighth phospholipid (PL8) at the donor-binding site in Stt3, to be described below; each of these lipid molecules has well-defined electron densities (Fig. 3c, d, Extended Data Fig. 3). PL1–PL3 fill a 15 Å gap at the interface between Stt3 and Ost1–Ost5. Their hydrophilic groups interact with Trp241, Gln250, Glu252, Asp301, Tyr303 and Tyr409 of Ost1, Arg112 and Asn113 of Stt3, and Tyr85 of Ost5. The lipid tails contact hydrophobic residues in TMH1–TMH2 of Stt3 and TMH2 of Ost5 (Fig. 3c). PL4–PL5 mediate the interaction between Stt3 and Ost2–Swp1 (Fig. 3d). Their tails interact with hydrophobic residues of TMH3 of Ost2, TMH2–TMH3 of Swp1, TMH5 and EL1 of Stt3, whereas the phosphate groups directly hydrogen bond to the side chains of Asn380, Asp381 and Arg385 of Wbp1. PL6–PL7 stabilize the interface of Stt3 and Ost2–Swp1. Many of these phospholipid-interacting residues are conserved (Extended Data Figs 7–9). Lipids are known to have important roles in membrane enzyme complexes^{34,35} and can mediate transient multimerization of membrane proteins³⁶. A published cryo-EM structure showed that lipids are involved in the assembly of the heterotetrameric γ -secretase³⁷.

Structures of the noncatalytic subunits

In the Ost1–Ost5 subcomplex, the Ost1 contains two luminal domains, NTD1 and NTD2; both are composed of a larger seven-stranded β -sheet with a smaller, four-stranded β -sheet attached at each end, and on the same face of, the larger sheet. NTD1 and NTD2 are superimposable with a root mean square deviation (r.m.s.d.) value of 2.8 Å, despite their 9% sequence identity (Fig. 4a, b). This fold is similar to the noncatalytic domain of leukotriene A4 aminopeptidase (also known as leukotriene A4 hydrolase) (r.m.s.d. value of 2.9 Å). Because Ost1 can bind the glycosylated, but not the unglycosylated, sequon²², NTD1 and NTD2 may function to prevent the glycosylated peptide from sliding back to the catalytic site. Perhaps related to this function, the Ost1 NTD2 has an extra conserved hydrophobic motif (relative to NTD1) that specifically binds to the Stt3 catalytic domain (Fig. 4b, Extended Data Fig. 8). Ost5 seems to be an accessory factor of Ost1, as its two TMHs pack against the sole TMH of Ost1, and the N-terminal luminal 20 residues of Ost5 latches onto the Ost1 NTD2, positioning Ost1 NTD2 for an interaction with Stt3 (Figs 3, 4).

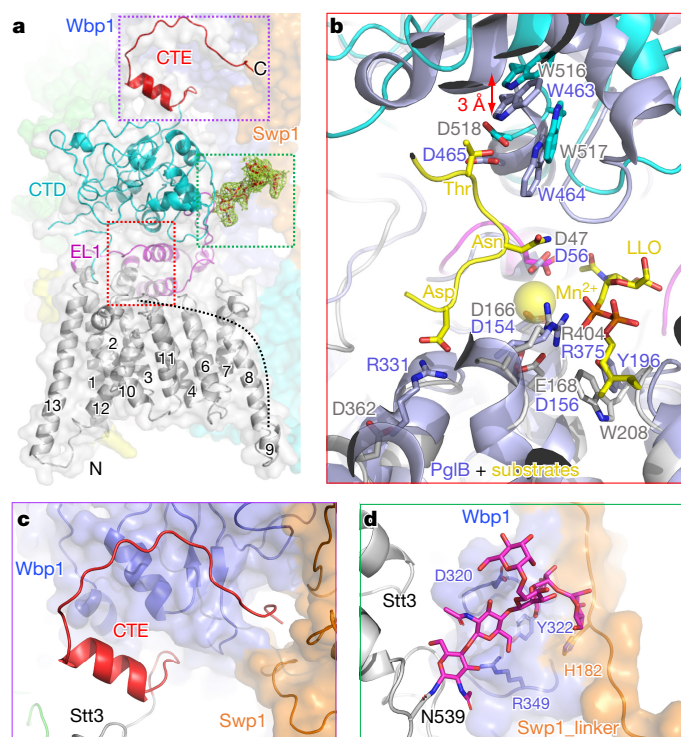


Figure 2 | The atomic structure of Stt3. **a**, Stt3 is shown as a cartoon. TMHs are in grey, CTD in cyan, EL1 in magenta, and CTE in red. The missing TMH9 is shown as a dotted black line. The green mesh is the *N*-glycan density of Asn539. The active site is highlighted by a dotted red square. The magenta dotted rectangle marks the Stt3 CTE interacting with Wbp1 and Swp1, and the green dotted square marks the *N*-glycan interacting with Wbp1 and Swp1. **b–d**, Enlargements of the dotted boxes in **a**. The active site of Stt3 (**b**) is superimposed with the PglB structure (PDB code 5OGL, light blue) in complex with Mn^{2+} , a peptide (DQNATF), and LLO analogue ((ωZZZ)-PPC-GlcNAc in yellow sticks). The red arrow indicates an outward shift of the Stt3 luminal domain relative to that of PglB.

The Ost2–Swp1–Wbp1 subcomplex contains three soluble domains: the NTD and middle domain of Wbp1 and NTD of Swp1 (Fig. 4c). The NTD and linker loop of Swp1 interact extensively with the NTD and middle domain of Wbp1. In the membrane region, four TMHs of Ost2 and three TMHs of Swp1 surround the sole TMH of Wbp1 (Fig. 4c). The functions of these domains are largely unknown, except that Wbp1 contains a GIFT domain (for the bacterial gliding protein GldD, intraflagellar transport (IFT)) and therefore may be involved in LLO binding³⁸. The Wbp1 NTD is superimposable with the NTD of IFT52 (Fig. 4d; r.m.s.d. of 2.9 Å), and that the middle domain of Wbp1 is similar to amylase domain N (Fig. 4e; r.m.s.d. of 2.7 Å), which has a pullulan-binding site. The Swp1 NTD resembles myeloid differentiation factor 1 (MD-1), a lipopolysaccharide (LPS)- and sugar-binding co-receptor of the RP105–MD-1 Toll-like receptor complex (Fig. 4f; r.m.s.d. of 3.5 Å). These results suggest that both Wbp1 and Swp1 are involved in recruiting LLO. Indeed, Wbp1 and its mammalian homologue OST48 were shown to cross-link with LLO^{21,39}.

In the Stt3–Ost4–Ost3 subcomplex, Ost4 stabilizes the Stt3 structure (Extended Data Fig. 5a, b). Ost3 has three TMHs that interact with TMH10, TMH11 and TMH13 of Stt3. TMH2 of Ost3 and TMH6 and TMH11 of Stt3 form a vertical groove that may function as LLO docking site (Fig. 5a, b). Consistent with this possibility, W208 at the upper half of the groove, close to the catalytic centre (Figs 2b, 5c), was previously shown to be lethal when mutated to alanine in Stt3⁴⁰. The corresponding Trp215Ala mutation in AglB reduced the activity¹¹, and the PglB equivalent residue Tyr196 directly interact with LLO¹². The lower half of this groove is lined with numerous hydrophobic residues and is occupied by phospholipid PL8 in our structure

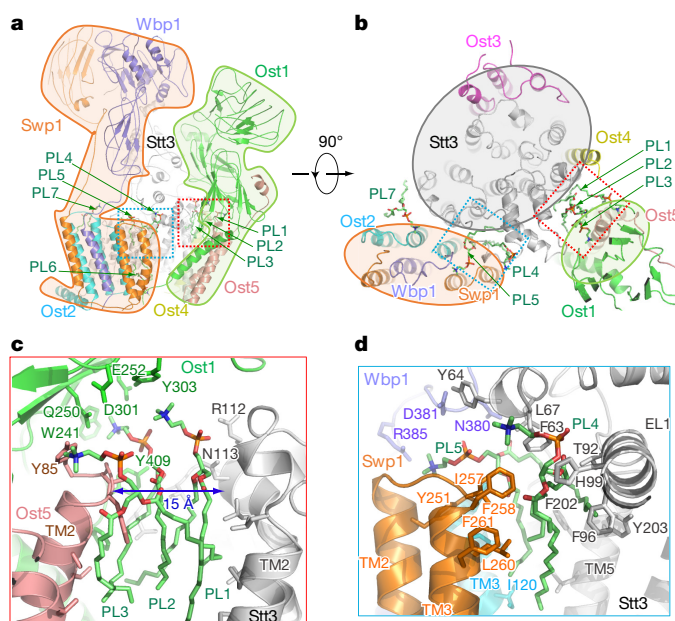


Figure 3 | Assembly of the OST complex. **a**, **b**, OST is shown as cartoon in a side view (**a**) and a top (luminal) view (**b**). Subcomplexes Ost1–Ost5, Ost2–Swp1–Wbp1 and Stt3–Ost3–Ost4 are highlighted by transparent shapes. The phospholipids PL1–PL7 are shown in green sticks. **c**, Close-up view of the red box in **a** and **b**. PL1–PL3 mediate the interactions between Stt3 and Ost1–Ost5, filling a 15 Å gap at the interface. **d**, Close-up view of the cyan box in **a** and **b**. PL4–PL6 mediate the interaction between Stt3 and Ost2–Swp1. PL6 in the lower leaflet of membrane is not visible here.

(Fig. 5c). Superposition of the structure of the PglB bound to the acceptor peptide and donor analogue with the Stt3 structure showed that the acceptor fits right into the putative active site and LLO analogue fits in potential LLO-docking vertical groove in Stt3 (Figs 2b, 5c).

The LLO entry route and lateral helix of Ost2

The EL5 loop of PglB is disordered in the absence of the donor, but becomes ordered when both donor and acceptor substrates are present^{12,13,41} (Fig. 5d). It was previously proposed that EL5 disordering allows the donor to diffuse under it into the catalytic site¹². Eukaryotic LLO is much larger, both in the transmembrane dolichol region and the luminal oligosaccharide region: the yeast dolichol contains 14–18 repeating isoprene units, twice that of bacterial LLO⁴², and the yeast oligosaccharide contains 14 sugars compared with the 7-sugar bacterial oligosaccharide. As such, the donor docking and recognition mechanisms of PglB and eukaryotic OST are probably different. Indeed, as described earlier, there is a large membrane-embedded pocket inside OST formed by TMH2 of Ost3 and TMH6, TMH8 and TMH11 of Stt3 (Fig. 5a, b). The disordered EL5 and TMH9 of Stt3 and TMH1 of Ost3 enlarge the donor-binding pocket. Furthermore, the Stt3 TMHs 8 and 9 slide towards the LLO-binding surface by about 25 Å relative to that of PglB (as described earlier) and form a lumenally unobstructed 10 Å-wide gap. The yeast oligosaccharide is probably too large to dive under the disordered EL5 to enter the catalytic site. We suggest that yeast LLO enters the catalytic site via the 10 Å gap between TMH8 and TMH9 in Stt3. This route has the added advantage of being closer to the potential LLO recruiters Wbp1 and Swp1 (Fig. 5a, b). Consistent with two distinct LLO access routes between OST and PglB, the PglB TMH8 and TMH9 are far from the LLO-binding site, tightly packed against other TMHs, and do not change their location whether the LLO is present or not (Fig. 5d, Extended Data Fig. 5c).

Ost2 has a cytoplasmic lateral α-helix (Figs 1c, 5a, b). The lateral α-helix stabilizes the Stt3 TMH8–TMH9 hairpin, and interacts with Swp1 TMH3 and WBP1 TMH1, both of which extend towards their respective luminal domains. Because the Stt3 TMH9 is largely

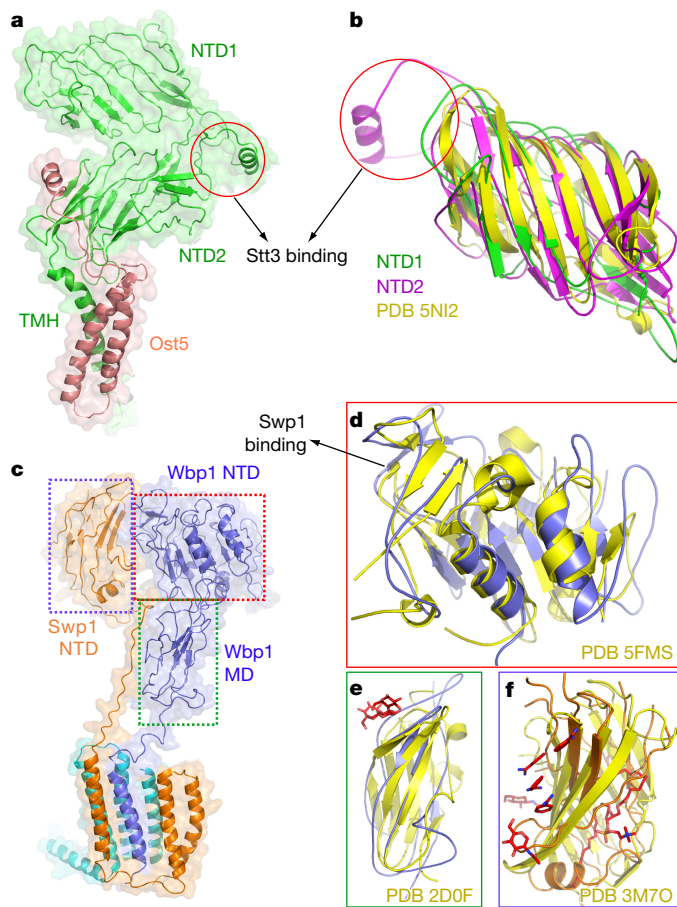


Figure 4 | The atomic structures of the noncatalytic subunits. **a**, Cartoon representation of subcomplex Ost1–Ost5. NTD1 and NTD2 of Ost1 are in green and magenta, respectively. Ost5 bridges Ost1 and Stt3. **b**, Superposition of NTD1 and NTD2 of Ost1 with the noncatalytic domain of human leukotriene A4 aminopeptidase. **c**, Cartoon representation of subcomplex Ost2–Wbp1–Swp1. The NTD and middle domain (MD) of Wbp1 and NTD of Swp1 are highlighted by three dotted squares. **d**, Superposition of the NTD of Wbp1 (blue) and the NTD of GIFT52 (yellow). **e**, Superposition of the middle domain of Wbp1 (blue) and the domain N of amylase (yellow). **f**, Superposition of the Swp1 NTD (orange) and the MD-1 (yellow). The red sticks in **e** and **f** are substrates in the homologue crystal structures.

disordered in the absence of donor, binding of dolichol will probably cause ordering of TMH9, which may be sensed by and communicated through the lateral α -helix to the luminal domains of Swp1 and Wbp1 via their respective TMH3 and TMH1. Because Swp1 and Wbp1 are involved in oligosaccharide binding, such a relay mechanism via the lateral α -helix may lead to coordination between the binding of dolichol in the membrane and OS binding in the lumen.

Ost3 mediates the OST–translocon interface

OST and the Sec61 complex simultaneously bind to ribosomes *in vitro*⁴³. Published cryo-electron tomographic studies showed the relative positions of mammalian translocons and OSTs bound to the ribosomes^{24,26}. However, the detailed interface—that is, which subunits mediate the OST–translocon interaction—was unclear. We docked the yeast OST atomic model into the cryo-electron tomogram and found that it fit well into the mammalian OST density except for two extra densities (Fig. 6a). These densities are likely to belong to the NTD of mammalian ribophorin II and CTD of ribophorin I, because only ribophorin I and II contain the extra sizable sequences (Extended Data Figs 8, 9). We further docked the crystal structure of a mammalian Sec61 into the tomogram and identified Ost3 as mediating the OST interaction

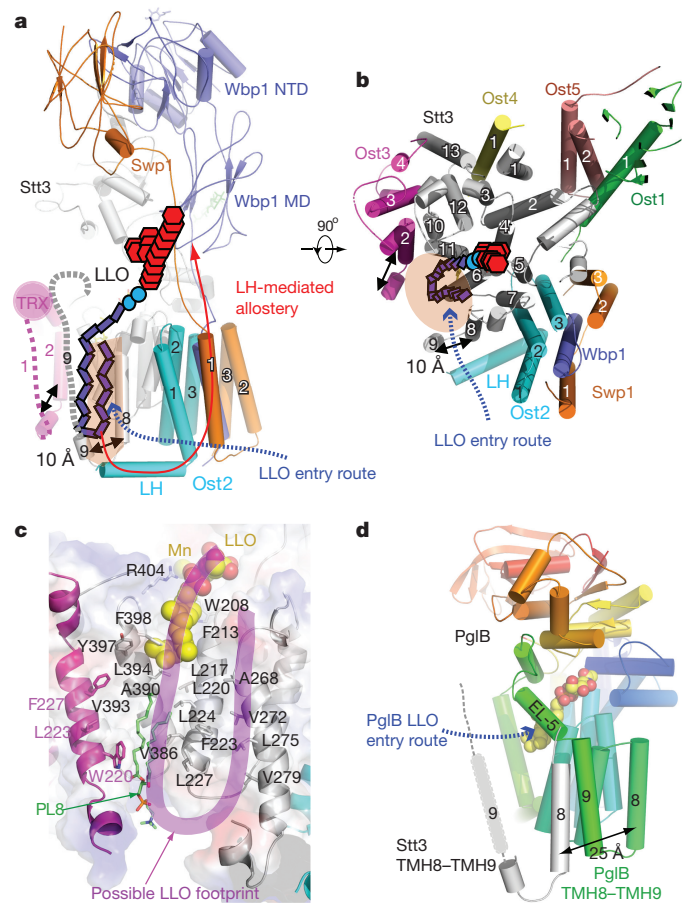


Figure 5 | A possible LLO entry route and allosteric coupling by Ost2 lateral α -helix. **a**, The unresolved Stt3 TMH9 (grey dotted line) and Ost3 TMH1 (magenta dotted line) are 10 Å away from the central body of OST, forming an enlarged LLO-binding site. The gap between TMH8 and TMH9 of Stt3 is probably the LLO entry gate. The red curve indicates a probable path for allostery mediated by the Ost2 lateral α -helix (LH). The purple rectangles, cyan circles and red hexagons represent dolichol, pyrophosphate and oligosaccharide, respectively. **b**, The top (luminal) view of the LLO-binding site. **c**, The LLO-binding hydrophobic surface in OST. PL8 is the eighth phospholipid at the substrate-binding surface. Mn^{2+} and the donor analogue in PglB structure are superimposed and shown in yellow spheres. **d**, Crystal structure of PglB (rainbow cartoon) in complex with LLO analogue (yellow spheres). Stt3 TMH8–TMH9 (grey cartoon) are superimposed.

with the translocon (Supplementary Video 3). Specifically, the Ost3 TMH3–TMH4 pack tightly with TMH1 of Sec61 α , TMH2 of Sec61 β and the sole TMH of Sec61 γ (Fig. 6b). This finding explains why the two yeast OST isoforms are defined by the presence of either Ost3 or Ost6: they interact with the Sec61 and Ssh1 translocons, respectively¹⁵. Our finding is also consistent with the finding that that DC2 and KCP2 mediate the interaction between the human OST complex and translocon⁴⁴.

On the basis of these observations, we proposed a pathway for the nascent peptide during the co-translational *N*-glycosylation (Fig. 6b). Specifically, the nascent peptide emerging from the Sec61 translocon is first captured by the luminal TRX of Ost3 and then is threaded through the Stt3 catalytic site for *N*-glycosylation. Finally, the glycosylated peptide is stabilized by the two luminal domains of Ost1, perhaps to prevent its backtracking.

Conclusion

Our cryo-EM study of the eukaryotic OST complex provides the atomic models of all eight component membrane proteins, reveals how these subunits assemble into a functional complex and suggests functions for

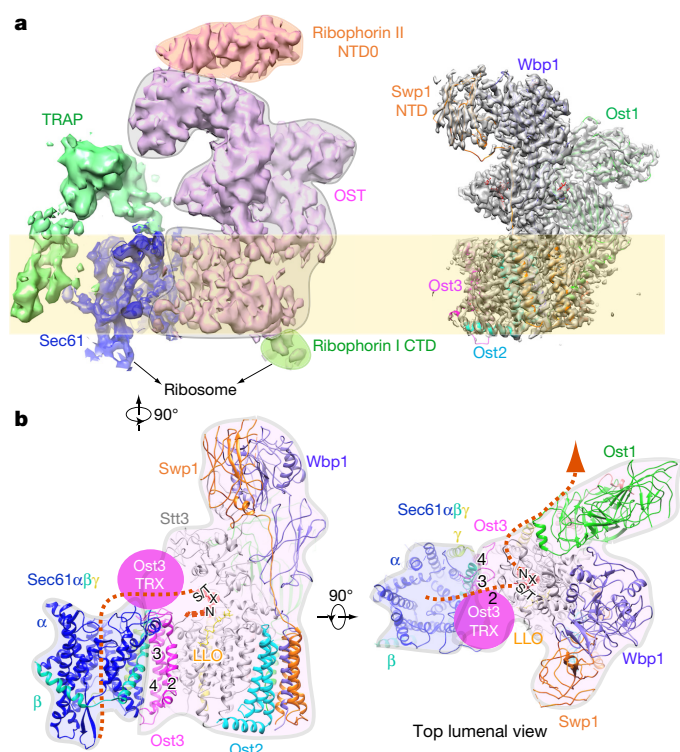


Figure 6 | A model of the OST-translocon super-complex.

a, Comparison of the yeast OST 3D map (right) with cryo-electron tomogram map of a mammalian OST (purple), which is in complex with translocon (blue), TRAP (translocon-associated protein) complex (green) and ribosome (not shown) (EMDB code EMD-3069). The mammalian OST complex has two extra domains: the cytosolic CTD in ribophorin I and luminal NTD0 of ribophorin II. **b**, A model of the OST-translocon super-complex, derived from docking OST structure and Sec61 structure (PDB code 3JC2). The dashed curve denotes the potential pathway for nascent peptide. See text for details.

many of the subunits. Given the crucial role of protein *N*-glycosylation in tumorigenesis and diagnosis^{45,46}, the structure may serve as a platform for the development of small-molecule inhibitors that target the *N*-glycosylation-related diseases in humans.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 December 2017; accepted 16 January 2018.

Published online 22 January 2018.

- Larkin, A. & Imperiali, B. The expanding horizons of asparagine-linked glycosylation. *Biochemistry* **50**, 4411–4426 (2011).
- Cherepanova, N., Shrimall, S. & Gilmore, R. N-linked glycosylation and homeostasis of the endoplasmic reticulum. *Curr. Opin. Cell Biol.* **41**, 57–65 (2016).
- Dempski, R. E. Jr & Imperiali, B. Oligosaccharyl transferase: gatekeeper to the secretory pathway. *Curr. Opin. Chem. Biol.* **6**, 844–850 (2002).
- Mohorko, E., Glockshuber, R. & Aebi, M. Oligosaccharyltransferase: the central enzyme of N-linked protein glycosylation. *J. Inher. Metab. Dis.* **34**, 869–878 (2011).
- Kornfeld, R. & Kornfeld, S. Assembly of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.* **54**, 631–664 (1985).
- Helenius, A. & Aebi, M. Roles of N-linked glycans in the endoplasmic reticulum. *Annu. Rev. Biochem.* **73**, 1019–1049 (2004).
- Apweiler, R., Hermjakob, H. & Sharon, N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim. Biophys. Acta* **1473**, 4–8 (1999).
- Hennet, T. & Cabalzar, J. Congenital disorders of glycosylation: a concise chart of glycosylation dysfunction. *Trends Biochem. Sci.* **40**, 377–384 (2015).
- Nothhaft, H. & Szymanski, C. M. Protein glycosylation in bacteria: sweeter than ever. *Nat. Rev. Microbiol.* **8**, 765–778 (2010).
- Matsumoto, S., Taguchi, Y., Shimada, A., Igura, M. & Kohda, D. Tethering an N-glycosylation sequon-containing peptide creates a catalytically competent oligosaccharyltransferase complex. *Biochemistry* **56**, 602–611 (2017).

- Matsumoto, S. *et al.* Crystal structures of an archaeal oligosaccharyltransferase provide insights into the catalytic cycle of N-linked protein glycosylation. *Proc. Natl Acad. Sci. USA* **110**, 17868–17873 (2013).
- Napiórkowska, M. *et al.* Molecular basis of lipid-linked oligosaccharide recognition and processing by bacterial oligosaccharyltransferase. *Nat. Struct. Mol. Biol.* **24**, 1100–1106 (2017).
- Lizak, C., Gerber, S., Numao, S., Aebi, M. & Locher, K. P. X-ray structure of a bacterial oligosaccharyltransferase. *Nature* **474**, 350–355 (2011).
- Kelleher, D. J. & Gilmore, R. An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology* **16**, 47R–62R (2006).
- Yan, A. & Lennarz, W. J. Two oligosaccharyl transferase complexes exist in yeast and associate with two different translocons. *Glycobiology* **15**, 1407–1415 (2005).
- Shrimall, S., Cherepanova, N. A. & Gilmore, R. Cotranslational and posttranslational N-glycosylation of proteins in the endoplasmic reticulum. *Semin. Cell Dev. Biol.* **41**, 71–78 (2015).
- Schulz, B. L. *et al.* Oxidoreductase activity of oligosaccharyltransferase subunits Ost3p and Ost6p defines site-specific glycosylation efficiency. *Proc. Natl Acad. Sci. USA* **106**, 11061–11066 (2009).
- Mohorko, E. *et al.* Structural basis of substrate specificity of human oligosaccharyl transferase subunit N33/Tusc3 and its role in regulating protein N-glycosylation. *Structure* **22**, 590–601 (2014).
- Gayen, S. & Kang, C. Solution structure of a human minimembrane protein Ost4, a subunit of the oligosaccharyltransferase complex. *Biochem. Biophys. Res. Commun.* **409**, 572–576 (2011).
- Zubkov, S., Lennarz, W. J. & Mohanty, S. Structural basis for the function of a minimembrane protein subunit of yeast oligosaccharyltransferase. *Proc. Natl Acad. Sci. USA* **101**, 3821–3826 (2004).
- Pathak, R., Hendrickson, T. L. & Imperiali, B. Sulfhydryl modification of the yeast Wbp1p inhibits oligosaccharyl transferase activity. *Biochemistry* **34**, 4179–4185 (1995).
- Yan, Q., Prestwich, G. D. & Lennarz, W. J. The Ost1p subunit of yeast oligosaccharyl transferase recognizes the peptide glycosylation site sequence, -Asn-X-Ser/Thr-. *J. Biol. Chem.* **274**, 5021–5025 (1999).
- Li, H., Chavan, M., Schindelin, H., Lennarz, W. J. & Li, H. Structure of the oligosaccharyl transferase complex at 12 Å resolution. *Structure* **16**, 432–440 (2008).
- Pfeffer, S. *et al.* Structure of the native Sec61 protein-conducting channel. *Nat. Commun.* **6**, 8403 (2015).
- Pfeffer, S. *et al.* Structure of the mammalian oligosaccharyl-transferase complex in the native ER protein translocon. *Nat. Commun.* **5**, 3072 (2014).
- Pfeffer, S. *et al.* Dissecting the molecular organization of the translocon-associated protein complex. *Nat. Commun.* **8**, 14516 (2017).
- Lara, P. *et al.* Refined topology model of the STT3/Stt3 protein subunit of the oligosaccharyltransferase complex. *J. Biol. Chem.* **292**, 11349–11360 (2017).
- Igura, M. *et al.* Structure-guided identification of a new catalytic motif of oligosaccharyltransferase. *EMBO J.* **27**, 234–243 (2008).
- Huang, C., Bhaskaran, R. & Mohanty, S. Eukaryotic N-glycosylation occurs via the membrane-anchored C-terminal domain of the Stt3p subunit of oligosaccharyltransferase. *J. Biol. Chem.* **287**, 32450–32458 (2012).
- Kowarik, M. *et al.* Definition of the bacterial N-glycosylation site consensus sequence. *EMBO J.* **25**, 1957–1966 (2006).
- Shrimall, S., Trueman, S. F. & Gilmore, R. Extreme C-terminal sites are posttranslationally glycosylated by the STT3B isoform of the OST. *J. Cell Biol.* **201**, 81–95 (2013).
- Li, G., Yan, Q., Nita-Lazar, A., Haltiwanger, R. S. & Lennarz, W. J. Studies on the N-glycosylation of the subunits of oligosaccharyl transferase in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **280**, 1864–1871 (2005).
- Yan, A., Ahmed, E., Yan, Q. & Lennarz, W. J. New findings on interactions among the yeast oligosaccharyl transferase subunits using a chemical cross-linker. *J. Biol. Chem.* **278**, 33078–33087 (2003).
- Palsdottir, H. & Hunte, C. Lipids in membrane protein structures. *Biochim. Biophys. Acta* **1666**, 2–18 (2004).
- Hunte, C. & Richers, S. Lipids and membrane protein structures. *Curr. Opin. Struct. Biol.* **18**, 406–411 (2008).
- Govaerts, C. Lipids can make them stick together. *Trends Biochem. Sci.* **42**, 329–330 (2017).
- Bai, X. C. *et al.* An atomic structure of human γ -secretase. *Nature* **525**, 212–217 (2015).
- Beatson, S. & Ponting, C. P. GIFT domains: linking eukaryotic intraflagellar transport and glycosylation to bacterial gliding. *Trends Biochem. Sci.* **29**, 396–399 (2004).
- Bause, E., Wesemann, M., Bartoschek, A. & Breuer, W. Epoxyethylglycyl peptides as inhibitors of oligosaccharyltransferase: double-labelling of the active site. *Biochem. J.* **322**, 95–102 (1997).
- Chavan, M., Rekowicz, M. & Lennarz, W. Insight into functional aspects of Stt3p, a subunit of the oligosaccharyl transferase. Evidence for interaction of the N-terminal domain of Stt3p with the protein kinase C cascade. *J. Biol. Chem.* **278**, 51441–51447 (2003).
- Shrimall, S., Cherepanova, N. A. & Gilmore, R. One flexible loop in OST lassos both substrates. *Nat. Struct. Mol. Biol.* **24**, 1009–1010 (2017).
- Kern, N. R. *et al.* Lipid-linked oligosaccharides in membranes sample conformations that facilitate binding to oligosaccharyltransferase. *Biophys. J.* **107**, 1885–1895 (2014).

43. Harada, Y., Li, H., Li, H. & Lennarz, W. J. Oligosaccharyltransferase directly binds to ribosome at a location near the translocon-binding site. *Proc. Natl Acad. Sci. USA* **106**, 6945–6949 (2009).
44. Shrima, S., Cherepanova, N. A. & Gilmore, R. DC2 and KCP2 mediate the interaction between the oligosaccharyltransferase and the ER translocon. *J. Cell Biol.* **216**, 3625–3638 (2017).
45. Hakomori, S. Glycosylation defining cancer malignancy: new wine in an old bottle. *Proc. Natl Acad. Sci. USA* **99**, 10231–10233 (2002).
46. Dube, D. H. & Bertozzi, C. R. Glycans in cancer and inflammation—potential for therapeutics and diagnostics. *Nat. Rev. Drug Discov.* **4**, 477–488 (2005).

Supplementary Information is available in the online version of the paper.

Acknowledgements Cryo-EM images were collected in the David Van Andel Advanced Cryo-Electron Microscopy Suite at Van Andel Research Institute. We thank Y. Harada for advice on yeast genetics and D. Nadziejka for

proofreading. This work was partially supported by Van Andel Research Institute (to H.L.) and the US National Institutes of Health (GM111742 to H.L.).

Author Contributions L.B. and H.L. designed the project. L.B. and A.K. purified proteins. L.B., T.W. and G.Z. collected cryo-EM data. L.B. processed data. L.B. and H.L. analysed the data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to H.L. (huilin.li@vai.org).

Reviewer Information *Nature* thanks S. Withers and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Purification of the OST complex. The OST complex was purified from the yeast strain LY500 as previously described²³. In brief, cells were lysed by French press at 103,000 kPa, and microsomes were collected by centrifugation. The membrane pellet was then resuspended and dissolved in buffer containing 10% glycerol, 20 mM Tris-HCl (pH 7.4), 1.5% digitonin, 0.5 M NaCl, 1 mM MgCl₂, 1 mM MnCl₂, 1 mM EDTA and 1 mM phenylmethylsulfonyl fluoride. After incubation for 30 min, the mixture was centrifuged for 30 min at 120,000 g, and the clarified supernatant was mixed with pre-washed anti-Flag (M2) affinity gel at 4 °C overnight with shaking. The affinity gel was collected by centrifugation and washed three times in buffer A containing 0.2% digitonin, 150 mM NaCl, 20 mM Tris-HCl (pH 7.4), 1 mM MgCl₂, 1 mM MnCl₂. Finally, the OST complex was eluted with buffer A containing 0.15 mg ml⁻¹ 3×Flag peptide, and further purified in a Superose 6 10/300 gel filtration column in buffer A. The quality of the final sample was confirmed by SDS-PAGE gel. To examine the presence of Ost3 and Ost6 in the purified complex, three bands around 30 kDa in the SDS-PAGE were cut out and identified by tryptic digestion and mass spectrometry.

Cryo-electron microscopy. Aliquots of 3 µl of purified OST at a concentration of 5 mg ml⁻¹ were placed on glow-discharged holey carbon grids (Quantifoil Au R1.2/1.3, 300 mesh) and flash-frozen in liquid ethane using an FEI VitroBot Mark IV. The grids were then loaded into an FEI Titan Krios electron microscope operated at a high tension of 300 kV and images were collected automatically with EPU (FEI) at a nominal magnification of 130,000× in EFTEM mode and a pixel size of 1.088 Å per pixel with defocus values from -1.5 to -2.5 µm. A Gatan K2 Summit direct electron detector was used under super resolution counting mode for image recording. A BioQuantum energy filter installed in front of the K2 detector was operated in zero-energy-loss mode with an energy slit width of 20 eV. The dose rate was 10 electrons per Å² per second, and the total exposure time was 6 s. The total dose was divided into a 30-frame movie so each frame was exposed for 0.2 s.

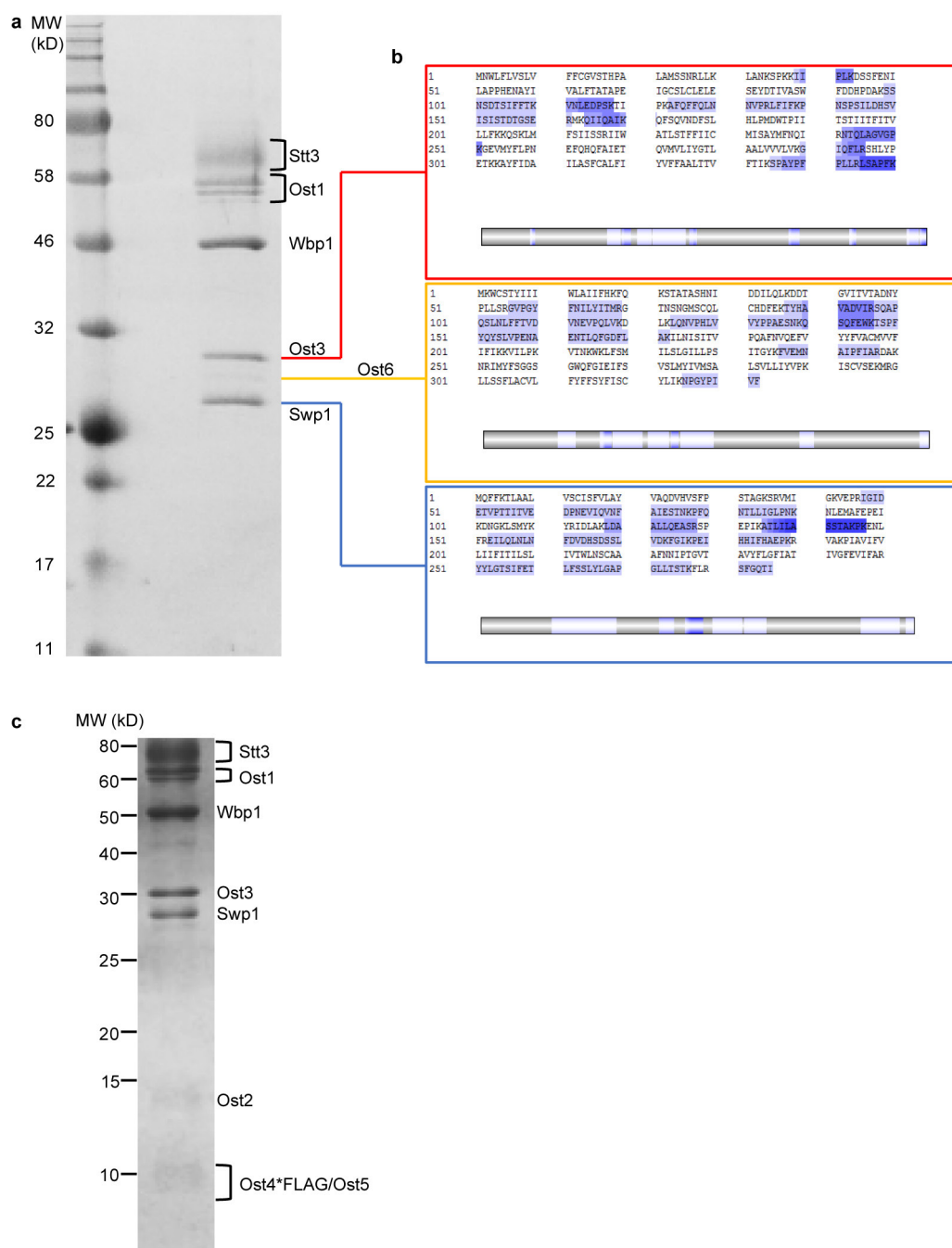
Image processing. About 4,000 raw movie micrographs were collected and motion-corrected using the program MotionCorr 2.0⁴⁷. Contrast transfer function parameters of each aligned micrograph were calculated using GCTF⁴⁸. All the remaining steps, including particle auto selection, 2D classification, 3D classification, 3D refinement and density map post-processing, were performed using RELION 2.0⁴⁹. Templates for automatic picking were generated from 2D averages of about 10,000 manually picked particles. A total of 823,255 particles were picked automatically. 2D classification was then performed and particles in the classes with features unrecognizable by visual inspection were removed. A total of 369,452 particles was used for further 3D classification, and 282,202 particles were selected for further 3D refinement and post-processing, resulting in the 3.5 Å 3D density map. The resolution of the map was estimated by the gold-standard Fourier shell correlation at a correlation cut-off value of 0.143.

Structural modelling, refinement and validation. The initial models of Stt3, the soluble domain of Ost1, and the NTD of Wbp1 were generated, respectively, from the crystal structures of *Archaeoglobus fulgidus* oligosaccharyltransferase

(PDB code 3WAK), leukotriene A4 hydrolase (PDB code 5NI2), and IFT52 (PDB code 5FMS) using the online server SWISSMODEL (<https://swissmodel.expasy.org>). The model of Stt3 was split into a transmembrane domain and a periplasmic domain. These models were docked into the 3.5 Å electron microscopy map in COOT and Chimera^{50,51}. All other subunits of OST were manually built into the remaining density in the program COOT. Sequence assignment was guided by bulky residues such as Phe, Tyr, Trp and Arg. The entire OST model was then refined by rigid-body refinement of individual chains in the PHENIX program and subsequently was adjusted manually in COOT⁵². There were densities for eight lipid molecules, each with well-defined densities for a head group and two tails. However, the precise chemical nature of the head group is unclear owing to the limited resolution. We modelled all lipids as a phosphatidylcholine, which is the most common lipid (approximately 60% phospholipid) in the ER membrane. The final model was also cross-validated as described before⁵³. Using the PDB tools in Phenix, the coordinates of the final model was first randomly added 0.1 Å noise, and then on this noise-added model was performed one round of refinement against the first half-map (half1) that was produced during 3D refinement by RELION. We then correlated the refined model with the 3D maps of the two half-maps (half1 and half2) to produce two Fourier shell correlation (FSC) curves: FSC_{work} (model versus half1 map) and FSC_{free} (model versus half2 map). Additionally, we generated a third FSC curve using the final model and the final 3.5 Å-resolution density map produced from all particles. The general agreement of these curves was taken as an indication that the model was not over-fitted. Finally, the atomic model was validated using MolProbity⁵⁴. Structural figures were prepared in Chimera and PyMOL (<https://pymol.org/2/>).

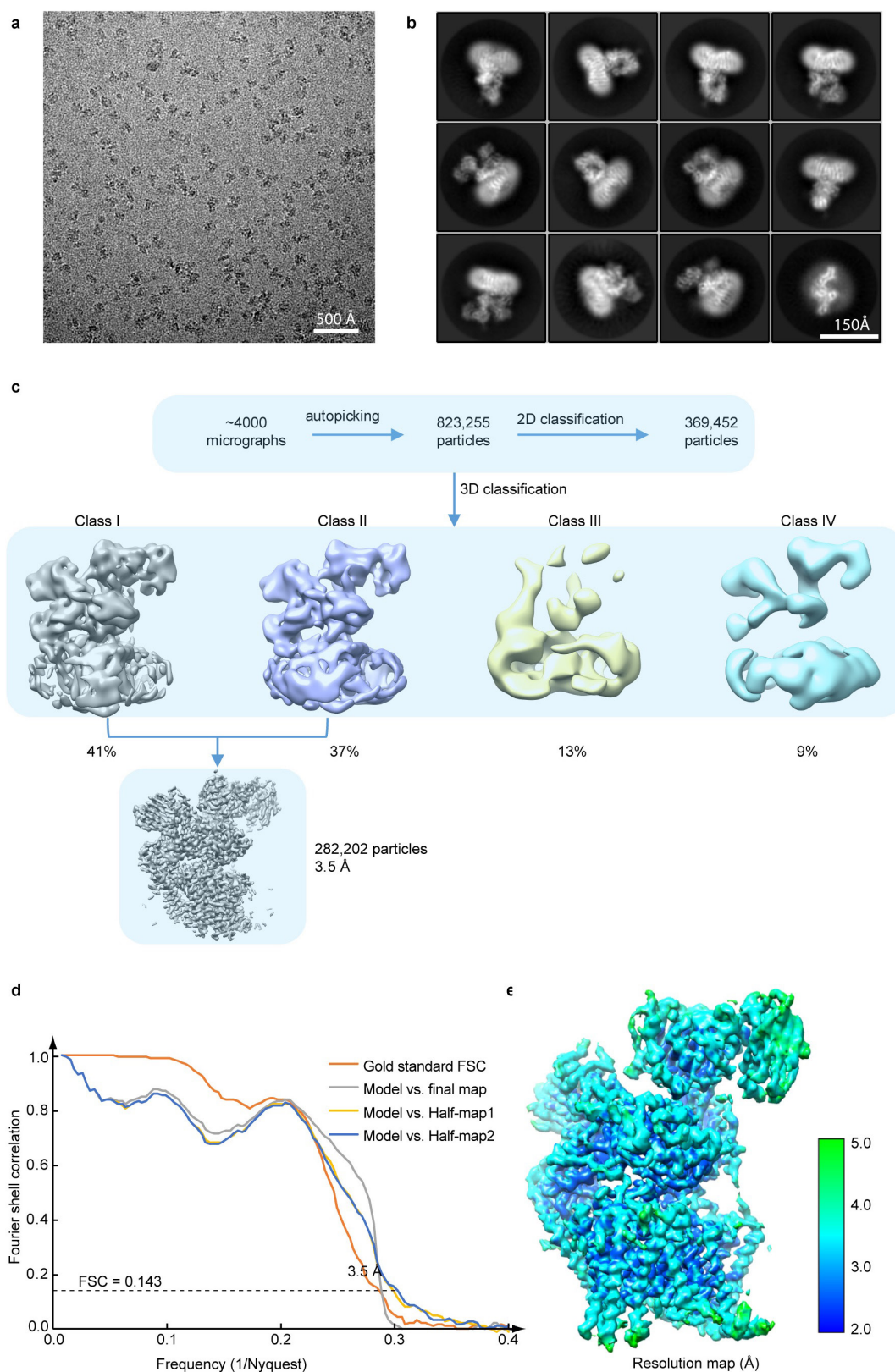
Data availability. The cryo-EM 3D map of the *S. cerevisiae* OST complex has been deposited at the Electron Microscopy Data Bank (EMDB) database with accession code EMD-7336. The corresponding atomic model was deposited at the RCSB Protein Data Bank (PDB) database with accession code 6C26.

47. Zheng, S. Q. *et al.* MotionCorr2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
48. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
49. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
50. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
51. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
52. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
53. Amunts, A. *et al.* Structure of the yeast mitochondrial large ribosomal subunit. *Science* **343**, 1485–1489 (2014).
54. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).



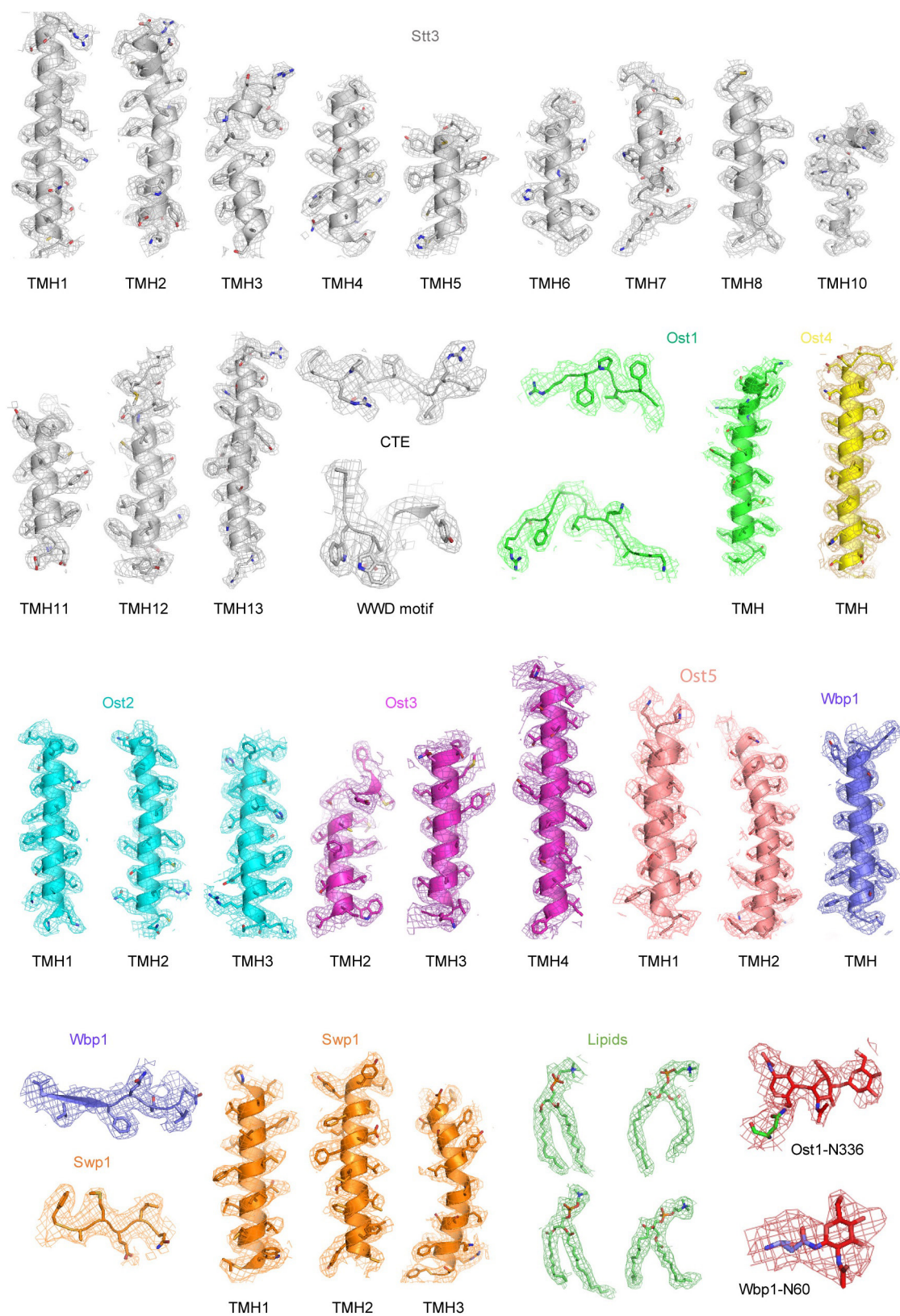
Extended Data Figure 1 | Identification of Ost3 and Ost6 by mass spectrometry. **a**, The Coomassie blue-stained SDS-PAGE gel of the purified OST complex. The small subunits Ost2, Ost4-Flag and Ost5 were not visible in this 12% acrylamide SDS-PAGE gel because of their weak density. **b**, Sequence coverage of tryptic digestion mass spectrometry of three bands at around 30 kDa that are labelled as Ost3, Ost6 and Swp1.

The detected peptides are highlighted in blue. The lower bars under the sequences indicate matched peptides. Darker blue indicates more overlaps of peptides detected. **c**, Ost2, Ost4-Flag and Ost5 were seen in the 15% acrylamide SDS-PAGE gel that was run slower and stained longer. Experiments in **a** and **c** were repeated more than three times with similar results.

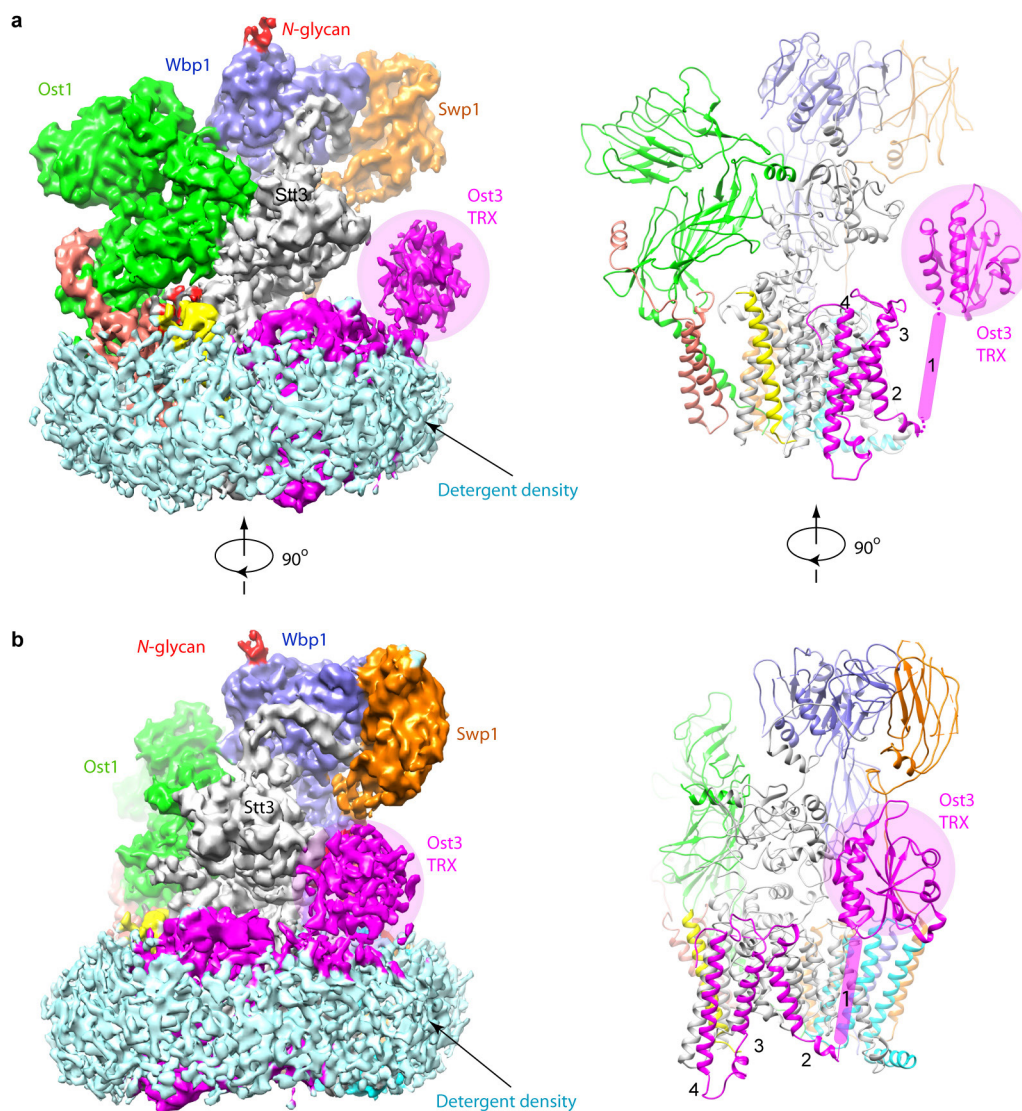


Extended Data Figure 2 | Single-particle cryo-EM analysis of the OST complex. **a**, A representative electron micrograph of the OST complex imaged in the Titan Krios with a K2 detector. About 4,000 similar micrographs were recorded. **b**, Selected reference-free 2D class averages. **c**, 2D and 3D image classification procedure. **d**, Gold-standard Fourier

correlation of two independent half maps, and the validation correlation curves of the atomic model by comparing the model with the final map or with the two half maps. **e**, Local resolution map of the OST complex structure.

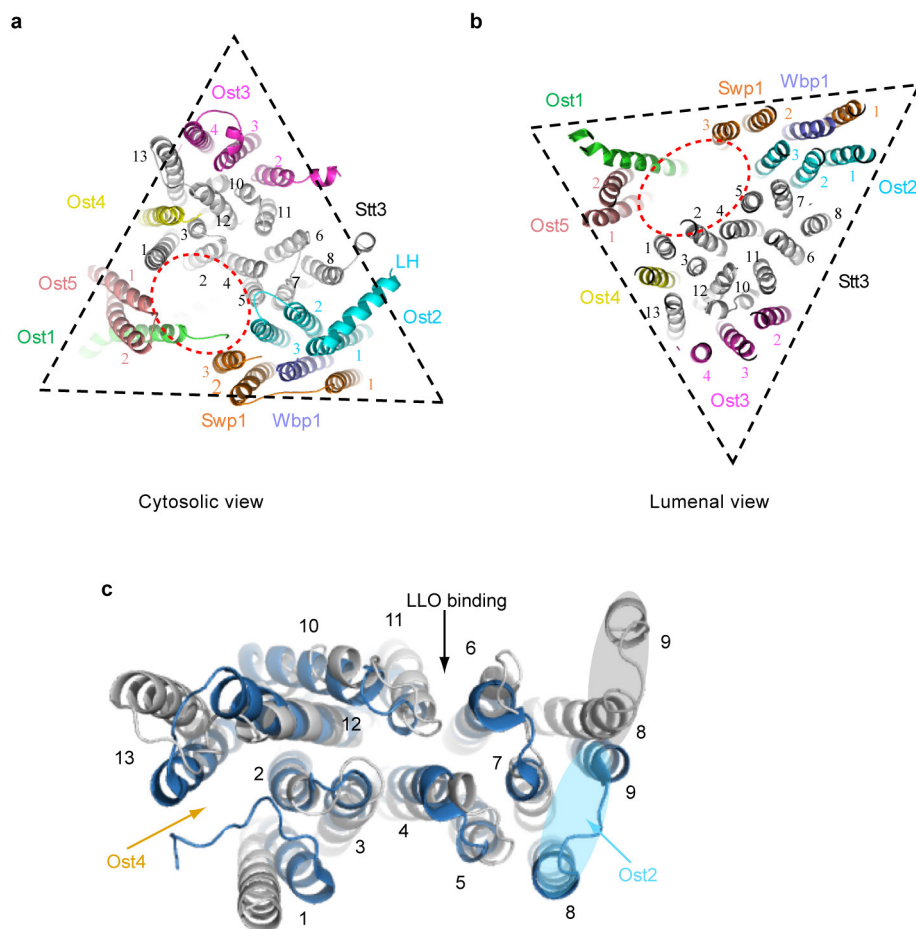


Extended Data Figure 3 | A gallery of selected regions in the OST structure, illustrating the fitting between the 3D density map and the atomic model. Selected regions in the structure of the OST complex include 26 TMHs, several regions in the luminal domains, four selected lipids and two N-glycans.



Extended Data Figure 4 | Electron density map of the TRX domain of Ost3. a, b, From 3D classification, one class (class I) contained stronger Ost3 TRX domain density than other classes. This map was further refined to 4.4 Å. Surface view of the map (left) and the corresponding cartoon view of the atomic model (right), coloured by subunit, are shown in two orthogonal side views. The N-terminal TRX domain of Ost3 is

highlighted by a magenta disk and is visible in this low-threshold display. The detergent densities that surround the transmembrane region of OST are visible at this threshold, and are coloured in cyan. The structure of the homologous Ost6 TRX (PDB code 3G7Y) is tentatively placed for the purpose of domain location.



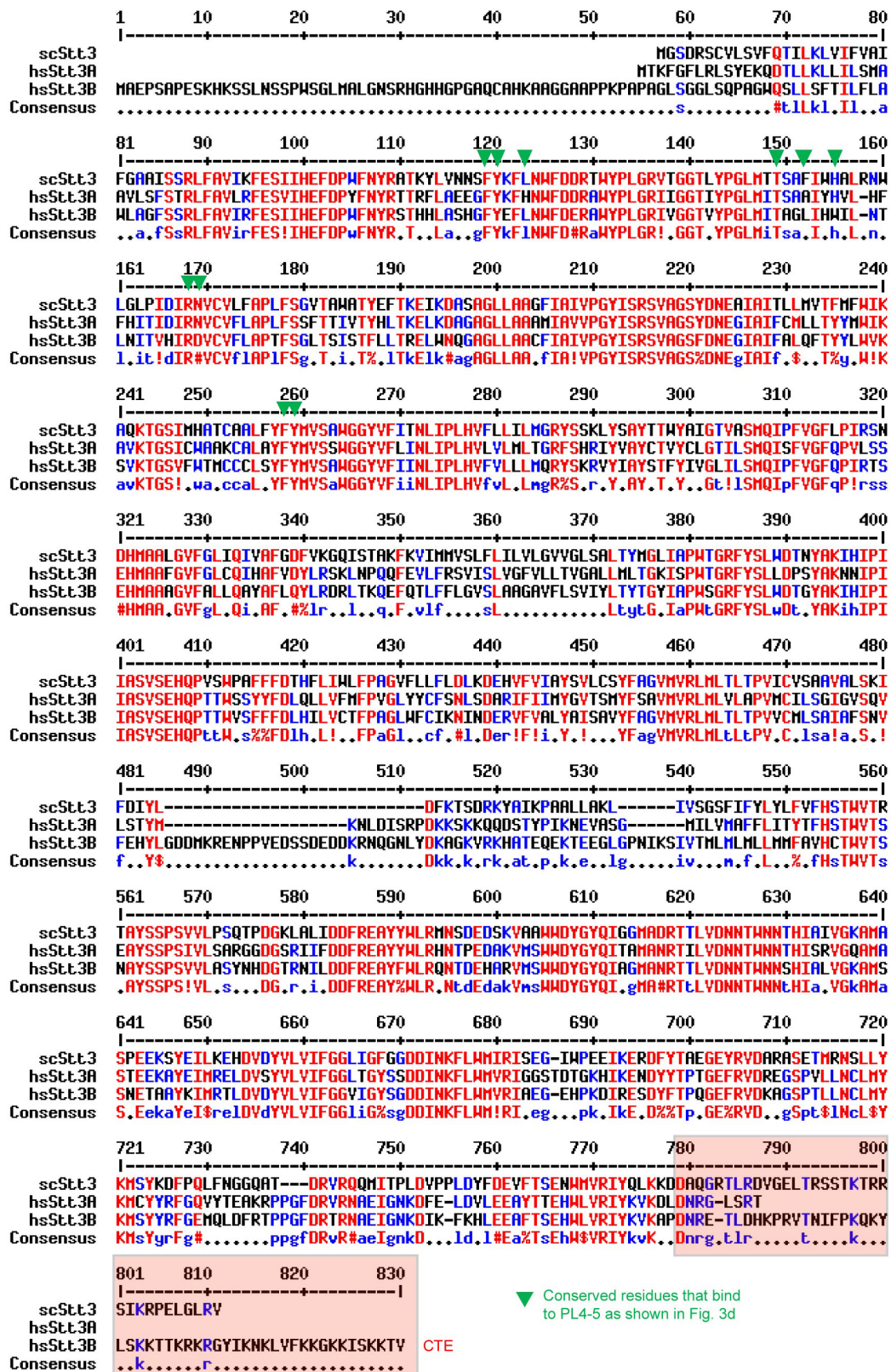
Extended Data Figure 5 | The transmembrane region of the OST complex. **a, b,** The TMHs of OST form a triangular shape and shown in cytoplasmic view (**a**) and luminal view (**b**). The catalytic subunit Stt3 is in the centre, surrounded by the other subunits. There is a sizable cavity in the centre (red dotted circle). **c,** Superposition of the transmembrane

region of Stt3 and PglB (PDB code 5OGL) viewed from the cytoplasmic side. The Stt3 TMH8–9 (light grey ellipse) moves towards the LLO binding surface relative to the TMH8–9 of PglB (light blue ellipse), creating space for the Ost2 TMHs. The Stt3 TMH1 and TMH13 also move apart, forming a space for the only TMH of Ost4.

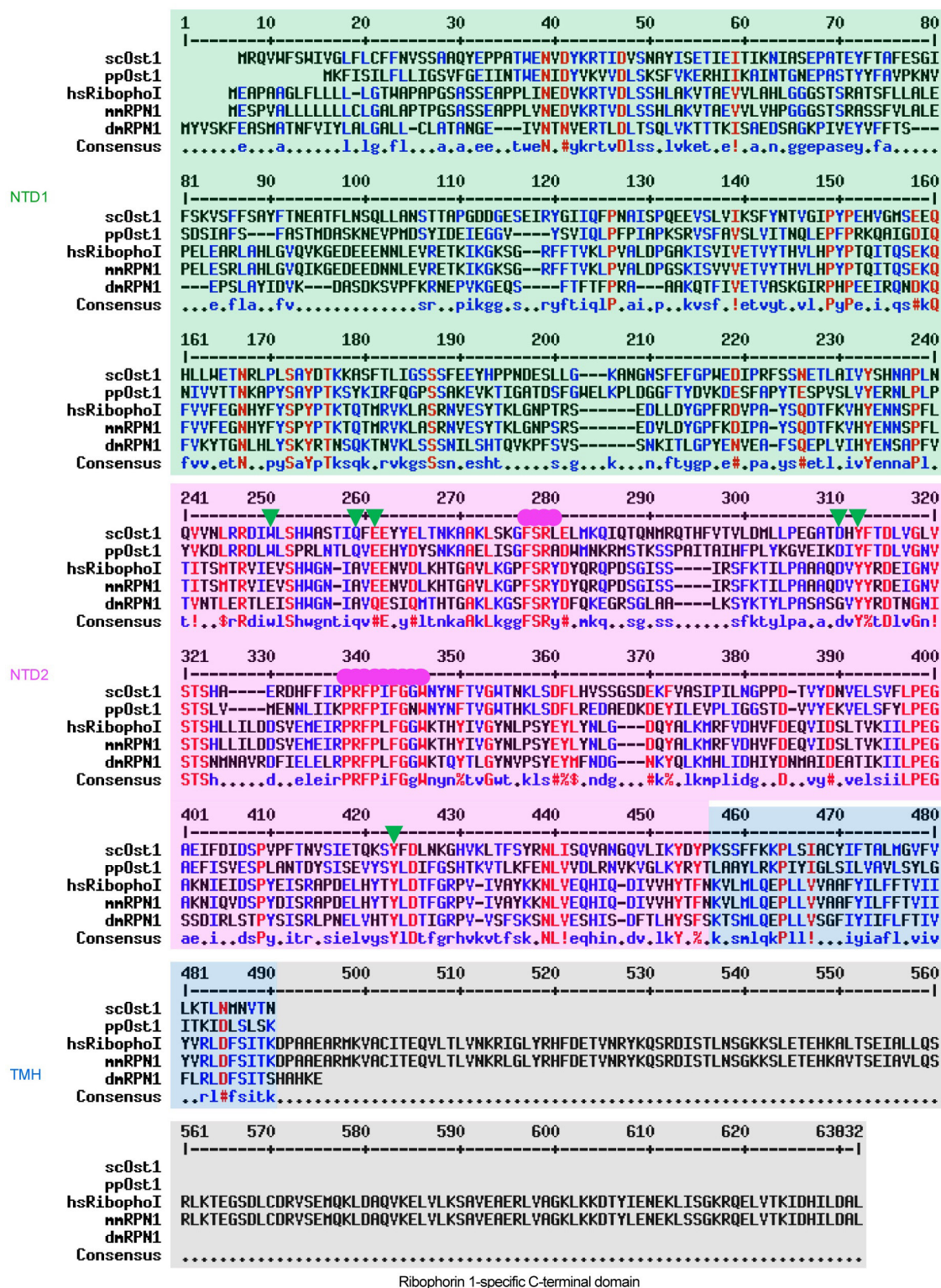
scStt3	1	MGSDRSCVLSVFQTIKLKLVIFVAIFGAAISSRLFAVIKFESIIHEFD	PWFNYRATKYLVNNSFYKFLNWFDD	72
PglB	1	MKLQQNFTDNNSIKYTCILIL-IAFAFSVLCRLYWVA-WASEFYEF--FFNDQ--LMITTNDGYAFAEGARD		66
cons		* :. . . . :*: * :. :*: * : * :*: **: : :...*. * * : *		
scStt3	73	RTWYPLGRVTGGTLYPGLMTTSAFIWHALRNWL--GLPIDIRNVCVLFAPLFSGVTAWATYEFTKEIKDASA		142
PglB	67	-----MIAGFHQPNDLSYFGSSLSTLTWLYSILPFSFESIILYMSAFFASLIVVPIILIAREYKLT		130
cons		: . * * . :. : * ** **: :. : : :*: :. : :*: * * :		
scStt3	143	GLLAAGFIAIVPGYISRSVAGSYDNEAIAITLLMVTFMFWIKAQKTGSIMHATCAALFYFYMVSAWGGYVFI		224
PglB	131	GFIAALLGSIANSYYNRMSGYYDTMLVLVLPMLILLTFIRLTINKDIFTLLSPVFIMIYLWWY-----		196
cons		*: **: :*. * .*:*: * *: :. :*. * *: :*: : . *: :. * : :		
scStt3	225	TNLIPLHVFLILMGRYSSKLYSAYTTWYAIGTVASMQUIPFVGFLPIRSNDHMAALGVFGLIQIVAFGDFVK		285
PglB	197	-----PSSYSLNFAMIGLFGLYTLVFHRK---		220
cons		* . :*: :*: **: * . .		
scStt3	286	GQISTAKFKVIMMVSLFLILVLGVVGLSALTYMGLIA-----PWTGRFYSLWDTNYAKIHI----		342
PglB	221	-----EKI-----FYLTIALMIIALSMLAWQYKLALIVLLFAIFAFKEEKINFYMIWALIFISILILHLS		280
cons		*: :*: :* :*: **: :* : . ** : * : * *		
scStt3	343	-----PIIASVSEHQPVSWPAFFFD--THFLIWLFPAGVFLLFLDL		381
PglB	281	GGLDPVLYQLKFYVFKASDVQNLKDAAFMYFNVNETIMEVNTIDPEVFMQRISSSVLVFILS--FIGFILLC		350
cons		: : * :. . *: : .*: :. : :*:		
scStt3	382	KDEHVFIAYSVLCSYFAGV--MVRMLTLTPVICVSAVALSKIFDIYDFKTSDRKYAIKPAALLAKLIV		451
PglB	351	KDHKSMLLALPMLALGFMALRAGLRFITYAVPVMALGFGYFLYAFFNF-L--EKKQIKLSLRNKNILLILIA		419
cons		**.: :*: :*. * :. :*: : .*: :. * :*: * :. : * :*: *		
scStt3	452	SGSFIFYLYLFVHSTWVTRTAYSSPSVVLPSQTPDGKLALIDDFREAYYWLRMNSDEDSKVAAWWDYGYQI		523
PglB	420	FFSISPALMHIYYYSSTVFTSY-----EASILND-----LKNKAQREDYVVAWWDYGYPI		470
cons		*: * : :. : **: : :*: * : :. : * .***** *		
scStt3	524	GGMADRTTLVDNNTW-----NNTHIAIVGKAM-----ASPEEKSYEILKEHDVDYV-----		569
PglB	471	RYYS DVKTLIDGGKHLGKDNFFSSFVLSKEQIPAAANMARLSVEYTEKSFKENYPDVLKAMVKDYNKTS AKDF		542
cons		: * .*: *... :. :* . . * * *: :*: **		
scStt3	570	-----LVIFGGLIGFGG-----		581
PglB	543	LESLNDKDFKFDTNKTRDVYIYMPYRMLRIMPVVAQFANTNPDNGEQEKS LFFSQANAIAQDKTTGSVMLDN		614
cons		:. *.. .*		
scStt3	582	--DDINKFLWM-----IRISEGIWP EEIKERDFYTAEGEYRVDARASETMRNSLLYKMSYKDFPQLFNGGQ		645
PglB	615	GVEIINDFRALKVEGASIPKAFVDIESITNGKFYYNE----IDSKA----QIYLLFLREYKSFVILDES LY		678
cons		: **.* : * :. : *.*: .** * :*: : **: .**.* * :.		
scStt3	646	ATDRVRQQMITPLDVPPLDYFDEVFTSENWMVRIYQLKKDDAQGRTL RDVGELTRSSTKTRRSIKRPELGLR		717
PglB	679	NSSYIQMFLLNQYD---QDLFEQIT--NDTRAKIYRLKR-----		712
cons		:. :. :. * * *: :. :*: *		
scStt3	718	<u>V</u>		
PglB	-			
cons				

Extended Data Figure 6 | Sequence alignments of *S. cerevisiae* Stt3 and *A. fulgidus* PglB. PglB does not have the CTE sequence (underscored) found in the yeast Stt3 and human STT3B. Several conserved residues in the active site are highlighted in red. R331 in the PglB, which stabilizes

the −2 position D of the acceptor peptide, is highlighted in blue. The asterisk indicates positions with identical residue, a colon indicates strong conservation and a full stop indicates weak conservation.

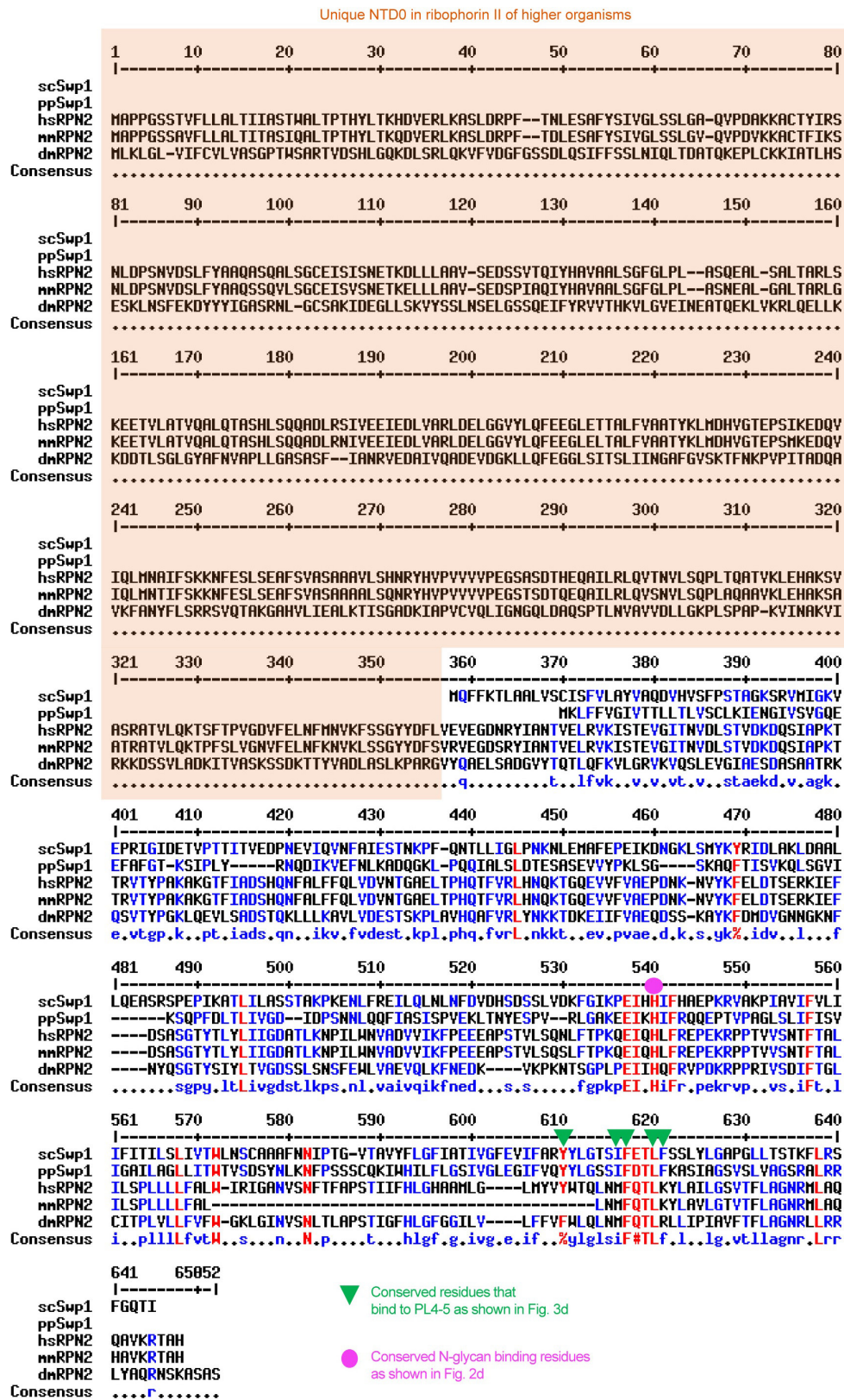


Extended Data Figure 7 | Sequence alignment of selected eukaryotic Stt3. The CTE of human STT3A is shorter than those of STT3B and yeast Stt3. hs, *Homo sapiens*; sc, *Saccharomyces cerevisiae*.



Extended Data Figure 8 | Sequence alignment of selected eukaryotic Ost1. An extra CTD in ribophorin I of complex eukaryotes such as flies, mice and humans is not present in the yeast proteins (shaded grey). NTD1

is shaded in light green, NTD2 in light magenta, TMH in light blue, and the CTD of ribophorin I of complex eukaryotes in light grey. dm, *Drosophila melanogaster*; pp, *Pichia pastoris*.



Extended Data Figure 9 | Sequence alignment of selected eukaryotic Swp1. Ribophorin II of complex eukaryotes has evolved an extra N-terminal domain (NTD0, shaded in light orange) in the lumen that is not present in the two yeast proteins.

Extended Data Table 1 | Cryo-EM data collection, refinement and validation statistics

	<i>S. cerevisiae</i> OST complex (EMD-7336) (PDB 6C26)
Data collection and processing	
Microscope	FEI Titan Krios
Voltage (kV)	300
Electron exposure (e ⁻ /Å ²)	60
Defocus range (μm)	1.5-2.5
Pixel size (Å)	1.088
Symmetry imposed	C1
Initial particle images (no.)	823255
Final particle images (no.)	282202
Map resolution (Å)	3.5
FSC threshold	0.143
Map resolution range (Å)	278.5-3.5
Refinement	
Map sharpening <i>B</i> factor (Å ²)	144.159
Model composition	
Non-hydrogen atoms	17202
Protein residues	2039
Lipids	8
N-glycans	3
R.m.s. deviations	
Bond lengths (Å)	0.01
Bond angles (°)	1.24
Validation	
MolProbity score	1.89
Clashscore	5.31
Poor rotamers (%)	1.15
Ramachandran plot	
Favored (%)	89.6
Allowed (%)	10.4
Disallowed (%)	0

Two chemically similar stellar overdensities on opposite sides of the plane of the Galactic disk

Maria Bergemann¹, Branimir Sesar², Judith G. Cohen³, Aldo M. Serenelli^{4,5}, Allyson Sheffield⁶, Ting S. Li⁷, Luca Casagrande^{8,9}, Kathryn V. Johnston¹⁰, Chervin F. P. Laporte¹⁰, Adrian M. Price-Whelan¹¹, Ralph Schönrich¹² & Andrew Gould^{1,13,14}

Our Galaxy is thought to have an active evolutionary history, dominated over the past ten billion years or so by star formation, the accretion of cold gas and, in particular, the merging of clumps of baryonic and dark matter^{1,2}. The stellar halo—the faint, roughly spherical component of the Galaxy—reveals rich ‘fossil’ evidence of these interactions, in the form of stellar streams, substructures and chemically distinct stellar components^{3–5}. The effects of interactions with dwarf galaxies on the content and morphology of the Galactic disk are still being explored. Recent studies have identified kinematically distinct stellar substructures and moving groups of stars in our Galaxy, which may have extragalactic origins^{6,7}. There is also mounting evidence that stellar overdensities (regions with greater-than-average stellar density) at the interface between the outer disk and the halo could have been caused by the interaction of a dwarf galaxy with the disk^{8–10}. Here we report a spectroscopic analysis of 14 stars from two stellar overdensities, each lying about five kiloparsecs above or below the Galactic plane—locations suggestive of an association with the stellar halo. We find that the chemical compositions of these two groups of stars are almost identical, both within and between these overdensities, and closely match the abundance patterns of stars

in the Galactic disk. We conclude that these stars came from the disk, and that the overdensities that they are part of were created by tidal interactions of the disk with passing or merging dwarf galaxies^{11,12}.

We present the spectroscopic analysis of 14 stars from two diffuse structures in the Milky Way halo, separated vertically by more than 10 kpc: the Triangulum–Andromeda (TriAnd) and A13 overdensities^{13–17}. TriAnd and A13 are located towards the Galactic anti-centre (the point, from the perspective of an observer on Earth, that is in the opposite direction to the Galactic centre), at Galactic latitudes (b) between -35° and -15° , and between $+25^\circ$ and $+40^\circ$, respectively. The age of stars in TriAnd is estimated from the colour–magnitude diagram to be 6–10 gigayears (Gyr)¹⁵. Studies of the motions of stars in these two structures revealed that they are kinematically associated¹⁷ and could be related to the Monoceros Ring—a ring-like stellar structure that twists around the Galaxy. However, the nature of the TriAnd and A13 structures remains hotly contested, with formation scenarios ranging from a disrupted dwarf galaxy to having their origin in the Galactic disk¹⁸.

We obtained high-resolution spectra of 14 stars using the Keck telescope and the Very Large Telescope (VLT; Extended Data Table 1). The

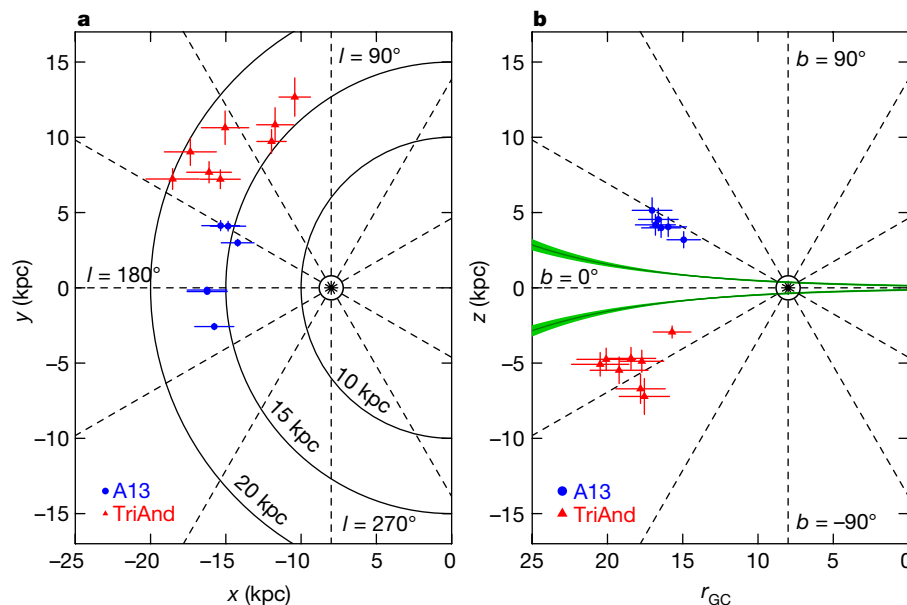


Figure 1 | Locations of the observed stars in Galactocentric Cartesian coordinates. **a, b,** The location of the 14 observed stars in the x – y plane (**a**) and the z – r_{GC} plane (**b**), where x , y and z are the directions of the Cartesian system of Galactic coordinates, and r_{GC} is the Galactocentric distance of the stars, defined as $\sqrt{x^2 + y^2}$, in kiloparsec (kpc). l and b are the Galactic longitude and latitude, respectively. The green curve in **b** indicates the flare profile of the Milky Way disk. The error bars represent the uncertainties of the distance measurements, which are estimated from the full posterior probability distributions to 1 standard deviation (s.d.).

¹Max Planck Institute for Astronomy, Königstuhl 17, 69117 Heidelberg, Germany. ²Deutsche Börse AG, Mergenthalerallee 61, 65760 Eschborn, Germany. ³Division of Physics, Mathematics, and Astronomy, California Institute of Technology, Pasadena, California 91125, USA. ⁴Institute of Space Sciences (ICE, CSIC), Carrer de Can Magrans, E-08193 Barcelona, Spain. ⁵Institut d'Estudis Espacials de Catalunya (IEEC), C/Gran Capità, 2-4, E-08034 Barcelona, Spain. ⁶Department of Natural Sciences, LaGuardia Community College, City University of New York, 31-10 Thomson Avenue, Long Island City, New York 11101, USA. ⁷Fermi National Accelerator Laboratory, PO Box 500, Batavia, Illinois 60510, USA. ⁸Research School of Astronomy and Astrophysics, Mount Stromlo Observatory, The Australian National University, Canberra, Australian Capital Territory 2611, Australia. ⁹ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia. ¹⁰Department of Astronomy, Columbia University, 550 West 120th Street, Mail Code 5246, New York, New York 10027, USA. ¹¹Department of Astrophysical Sciences, Princeton University, 4 Ivy Lane, Princeton, New Jersey 08544, USA. ¹²Rudolf-Peierls Centre for Theoretical Physics, University of Oxford, 1 Keble Road, Oxford OX1 3NP, UK. ¹³Korea Astronomy and Space Science Institute, Daejeon 34055, South Korea. ¹⁴Department of Astronomy, Ohio State University, 140 West 18th Avenue, Columbus, Ohio 43210, USA.

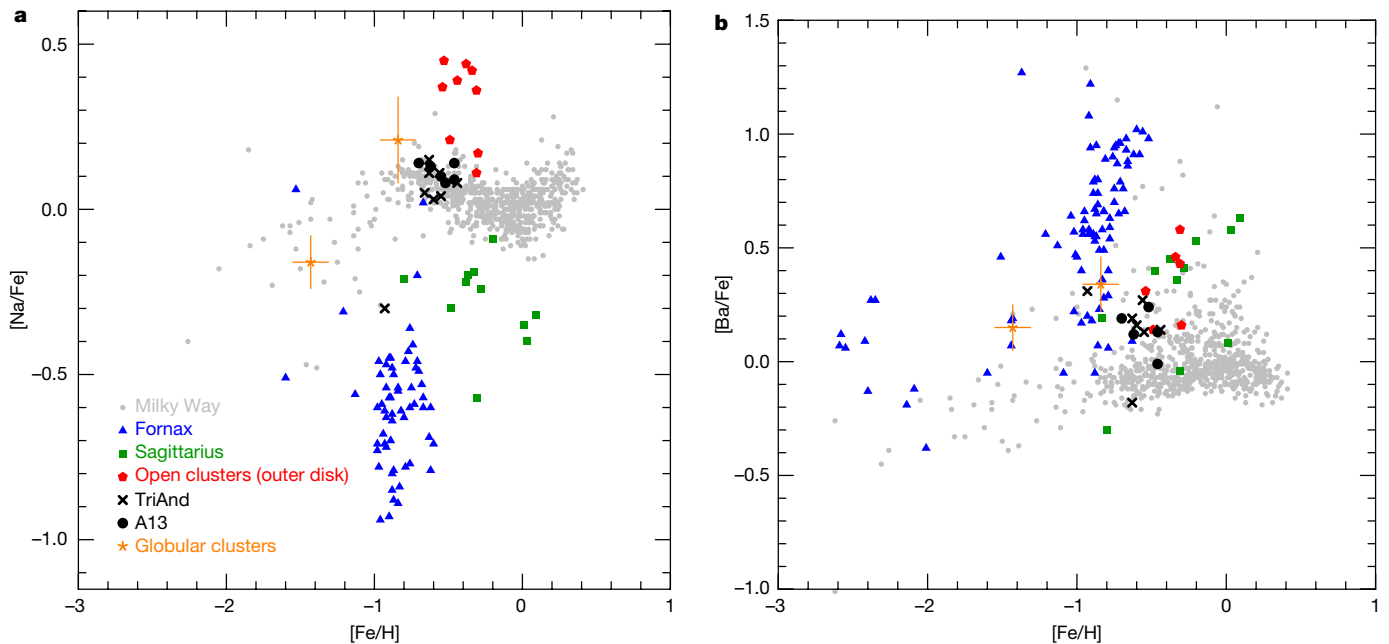


Figure 2 | Chemical abundances of the observed stars. **a, b,** Chemical abundance ratios $[\text{Na}/\text{Fe}]$ (**a**) and $[\text{Ba}/\text{Fe}]$ (**b**) versus metallicity ($[\text{Fe}/\text{H}]$) in: the TriAnd and A13 overdensities; the Milky Way disk and halo stars; the Fornax and Sagittarius dwarf spheroidal galaxies; open clusters in the Galactic outer disk; and globular clusters (with error bars reflecting

intracluster abundance variation, derived as the root-mean-square (r.m.s.) variance of the sample, with $N=13$ for the M3 cluster and $N=25$ for M71). Further details and references are provided in Methods, and source data are given in Extended Data Tables 2, 3. The typical uncertainty of the abundance measurements is 0.15 dex.

stars are confirmed members of the A13 and TriAnd overdensities on the basis of their radial velocities, proper motions and photometry. We determined fundamental atmospheric parameters of the stars, as well as chemical abundances for oxygen, sodium, magnesium, titanium, iron, barium and europium, by combining analysis of their colours with standard spectroscopic methods (Methods). We derived stellar distances from 2MASS photometry and spectroscopic gravities, which place the TriAnd stars at a Galactocentric distance of $r_{\text{GC}} = 18 \pm 2$ kpc (where 2 kpc is roughly one standard deviation, 1σ), about 5 kpc below the Galactic disk plane, and A13 stars at $r_{\text{GC}} = 16 \pm 1$ kpc, some 4 kpc above the plane (Fig. 1). The typical distance uncertainties are about 1–2 kpc (Extended Data Table 2). From the same spectra, we determine σ_{los} , the line-of-sight velocity dispersion of the sample of stars (see Methods), to be 27 km s^{-1} , markedly lower than that of the halo stars^{12,19}, which have σ_{los} values of about 100 km s^{-1} . The rotational velocity for the stars in the sample is $195 \pm 25 \text{ km s}^{-1}$, consistent with the circular velocity in the outer disk.

Our analysis shows that the abundance distribution in A13 and TriAnd is extremely compact, and that the spread is consistent with observational errors, which are about 0.15 dex. For the abundances, we use the notation $[A/B]$, which refers to the logarithm of the abundance ratio of the chemical element A to the element B, scaled to the solar value. Our results for the mean abundance ratios in the overdensities are: $\langle [\text{Fe}/\text{H}] \rangle = -0.59 \pm 0.12$ dex, $\langle [\text{O}/\text{Fe}] \rangle = 0.24 \pm 0.11$ dex; $\langle [\text{Na}/\text{Fe}] \rangle = 0.09 \pm 0.11$ dex, $\langle [\text{Mg}/\text{Fe}] \rangle = 0.20 \pm 0.03$ dex, $\langle [\text{Ti}/\text{Fe}] \rangle = 0.08 \pm 0.09$ dex, $\langle [\text{Ba}/\text{Fe}] \rangle = 0.14 \pm 0.13$ dex, and $\langle [\text{Eu}/\text{Fe}] \rangle = 0.20 \pm 0.16$ dex (Extended Data Table 3), where the spread is given by the sample standard deviation. We compare these abundance ratios with literature measurements of stars from the Galactic disk and halo, dwarf spheroidal galaxies and globular clusters, finding that our measurements are consistent with abundances in the ‘thin’ disk—the younger component of the Milky Way—but are inconsistent with all other stellar populations (Fig. 2 and Extended Data Fig. 1). All but one stars from the two overdensities lie directly on the metal-poor end of the thin-disk track, which represents stars in the outer disk of the Galaxy²⁰. The only TriAnd star with slightly lower metallicity, $[\text{Fe}/\text{H}] \approx -0.9$, resides on the canonical ‘thick-disk’ track which has

higher $[\text{Ba}/\text{Fe}]$ and lower $[\text{Na}/\text{Fe}]$ ratios. The overdensity abundances are also consistent with those derived from stars in open clusters^{21,22} and with Cepheids^{23,24} at comparable Galactocentric distances, in the r_{GC} range 10–20 kpc.

The similarity of abundances suggests that the TriAnd and A13 stars have a common origin. However, having their origin in a star cluster is very unlikely, as stellar clusters are compact in physical space, in sharp contrast to the overdensity stars. A tidal disruption of a globular cluster, as occurred for example with the Palomar 5 cluster²⁵, can cause individual stars to be strewn over large distances in one direction on the sky. However, unlike the overdensities, the tidal tails of Palomar 5 are thin in the directions transverse to the resulting stellar stream. Also, the A13 and TriAnd stars do not exhibit the anti-correlation between sodium and oxygen abundances that is found in almost all globular clusters in the Milky Way²⁶.

On the other hand, dwarf spheroidal galaxies—which can extend over several kiloparsecs, and can be disrupted to extend over tens of kiloparsecs—show a much larger scatter in abundance space (Fig. 2 and Extended Data Fig. 1), which is thought to be due to multiple generations of star formation²⁷. Also, in contrast to the stars in the overdensities, dwarf spheroidal galaxies are known to extend to very low, typically sub-solar, $[\text{O}/\text{Fe}]$, $[\text{Mg}/\text{Fe}]$ and $[\text{Na}/\text{Fe}]$ ratios at $[\text{Fe}/\text{H}]$ ratios of about -0.5 dex. Fornax, the dwarf galaxy that is closest to TriAnd in metallicity, has an $[\text{Na}/\text{Fe}]$ ratio of about -0.6 dex, which is almost one order of magnitude lower than the relative sodium abundance of the A13 overdensity stars. On the other hand, the relative barium abundance of the Fornax stars is a factor of ten higher than that of the A13 and TriAnd stars.

The premise that the origin of the A13 and TriAnd stars is in the Milky Way disk is strongly supported by the stellar chemical abundances and motions. The key challenge to this hypothesis is that the stars are located very far away from the disk plane and at large Galactocentric distances of more than 15 kpc. A plausible scenario that may explain our observations is related to a merger of a dwarf galaxy with the Milky Way disk. Simulations show that such mergers can trigger vertical oscillations and flaring in the pre-existing disk, which naturally explain the existence of stellar overdensities above and below the Galactic midplane^{9,28}. To test

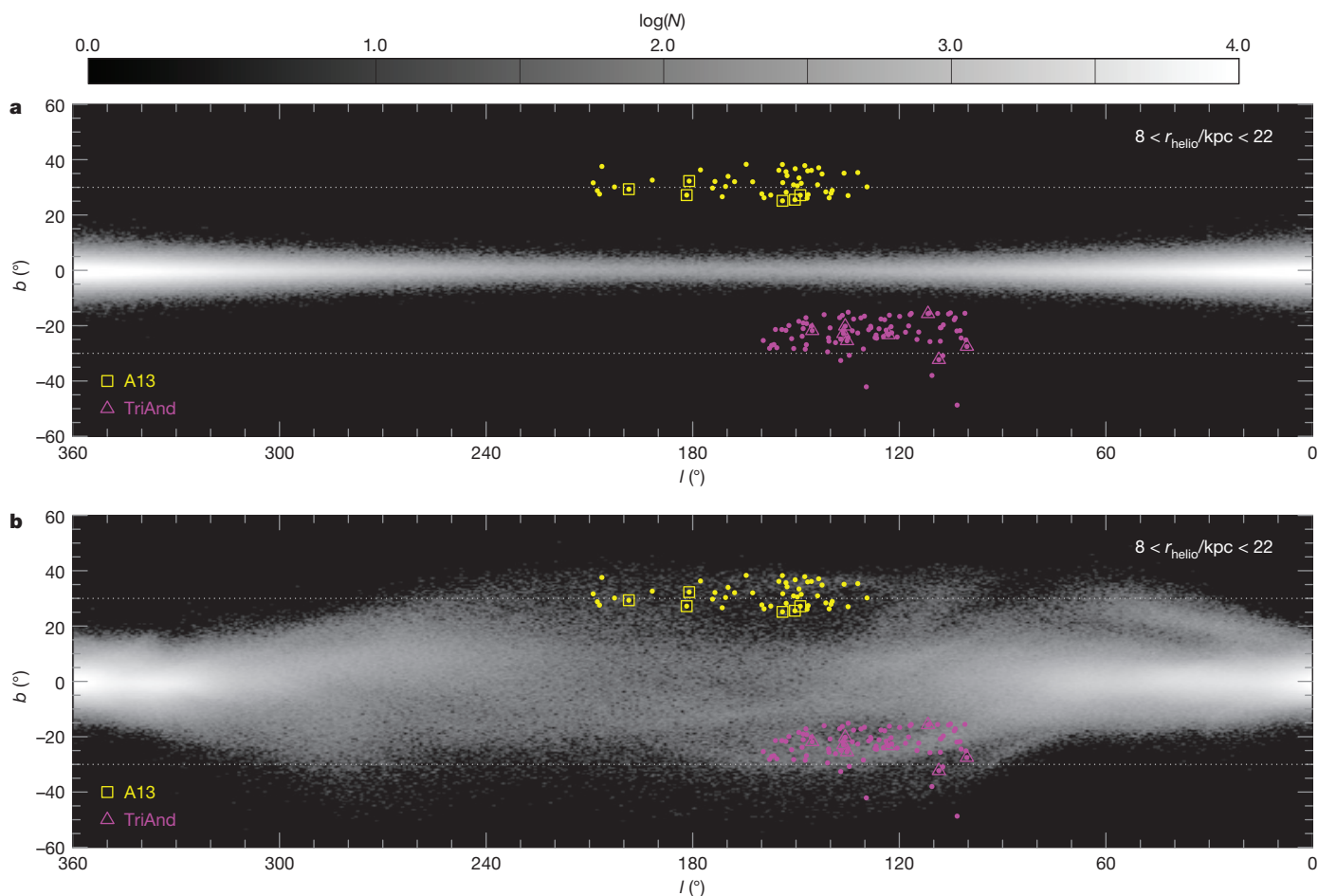


Figure 3 | Comparison of the positions of the observed stars with an N -body simulation. The locations of the observed A13 and TriAnd stars (yellow squares and purple triangles, respectively) are compared with two snapshots from an N -body model (represented in grey) that follows the interaction of the Sagittarius dwarf spheroidal galaxy with an initially

stable Galactic disk²⁹. **a**, Distribution of star particles at the beginning of the simulation. **b**, The final, present-day distribution, after 5.6 Gyr. Other confirmed members of the overdensities^{15,17} are also shown, as small yellow and magenta dots. r_{helio} , heliocentric distance.

this scenario, we compared the spatial locations of the TriAnd and A13 stars with the predictions of an N -body model, which follows the interaction of the Sagittarius dwarf spheroidal galaxy with an initially stable Galactic disk²⁹ (Fig. 3). In this model, the initial dark-halo mass for the Sagittarius progenitor was about $10^{11} M_{\odot}$ (where M_{\odot} is the mass of the Sun), which, after 5.57 Gyr of evolution, was stripped down to a bound mass of $3 \times 10^9 M_{\odot}$. Initially, the disk in the region of A13 and TriAnd, corresponding to heliocentric distances of 8–22 kpc, was confined between the latitudes of -5° and $+5^{\circ}$. However, during its interaction with the Milky Way, Sagittarius was able to excite the outer disk outwards as far as -30° to $+30^{\circ}$, kicking material well above the midplane out into the regions of our TriAnd and A13 stars. Therefore, the results of these simulations—which also suggest appropriate timescales for the interaction of the Sagittarius dwarf spheroidal galaxy with the Milky Way, varying from 5 Gyr to 9 Gyr—support our interpretation that these stars may have originated from the disk.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 August; accepted 11 December 2017.

Published online 26 February 2018

- Freeman, K. & Bland-Hawthorn, J. The new galaxy: signatures of its formation. *Annu. Rev. Astron. Astrophys.* **40**, 487–537 (2002).
- Bullock, J. S. & Johnston, K. V. Tracing galaxy formation with stellar halos. I. Methods. *Astrophys. J.* **635**, 931–949 (2005).

- Helmi, A., White, S. D. M., de Zeeuw, P. T. & Zhao, H. Debris streams in the solar neighbourhood as relicts from the formation of the Milky Way. *Nature* **402**, 53–55 (1999).
- Bell, E. F. *et al.* The accretion origin of the Milky Way's stellar halo. *Astrophys. J.* **680**, 295–311 (2008).
- Nissen, P. E. & Schuster, W. J. Two distinct halo populations in the solar neighborhood. Evidence from stellar abundance ratios and kinematics. *Astron. Astrophys.* **511**, L10 (2010).
- Helmi, A. *et al.* Pieces of the puzzle: ancient substructure in the Galactic disc. *Mon. Not. R. Astron. Soc.* **365**, 1309–1323 (2006).
- Stonkute, E., Tautvaišienė, G., Nordström, B. & Ženovienė, R. Stellar substructures in the solar neighbourhood. I. Kinematic group 3 in the Geneva-Copenhagen survey. *Astron. Astrophys.* **541**, A157 (2012).
- Xu, Y. *et al.* Rings and radial waves in the disk of the Milky Way. *Astrophys. J.* **801**, 105 (2015).
- Purcell, C. W., Bullock, J. S., Tollerud, E. J., Rocha, M. & Chakrabarti, S. The Sagittarius impact as an architect of spirality and outer rings in the Milky Way. *Nature* **477**, 301–303 (2011).
- Dorman, C. E. *et al.* A new approach to detailed structural decomposition from the SPLASH and PHAT surveys: kicked-up disk stars in the Andromeda galaxy? *Astrophys. J.* **779**, 103 (2013).
- Kazantzidis, S., Bullock, J. S., Zentner, A. R., Kravtsov, A. V. & Moustakas, L. A. Cold dark matter substructure and galactic disks. I. Morphological signatures of hierarchical satellite accretion. *Astrophys. J.* **688**, 254–276 (2008).
- Price-Whelan, A. M., Johnston, K. V., Sheffield, A. A., Laporte, C. F. P. & Sesar, B. A reinterpretation of the Triangulum-Andromeda stellar clouds: a population of halo stars kicked out of the Galactic disc. *Mon. Not. R. Astron. Soc.* **452**, 676–685 (2015).
- Rocha-Pinto, H. J., Majewski, S. R., Skrutskie, M. F., Crane, J. D. & Patterson, R. J. Exploring halo substructure with giant stars: a diffuse star cloud or tidal debris around the Milky Way in Triangulum-Andromeda. *Astrophys. J.* **615**, 732–737 (2004).

14. Martin, N. F., Ibata, R. A. & Irwin, M. Galactic halo stellar structures in the Triangulum-Andromeda region. *Astrophys. J.* **668**, L123–L126 (2007).
15. Sheffield, A. A. *et al.* Exploring halo substructure with giant stars. XIV. The nature of the Triangulum-Andromeda stellar features. *Astrophys. J.* **793**, 62 (2014).
16. Sharma, S. *et al.* Group finding in the stellar halo using M-giants in the Two Micron All Sky Survey: an extended view of the Pisces overdensity? *Astrophys. J.* **722**, 750–759 (2010).
17. Li, T. S. *et al.* Exploring halo substructure with giant stars. XV. Discovery of a connection between the Monoceros Ring and the Triangulum-Andromeda overdensity? *Astrophys. J.* **844**, 74 (2017).
18. Johnston, K. V. *et al.* Disk heating, galactoseismology, and the formation of stellar halos. *Galaxies* **5**, 3 (2017).
19. Xue, X. X. *et al.* The Milky Way's circular velocity curve to 60 kpc and an estimate of the dark matter halo mass from the kinematics of 2400 SDSS blue horizontal-branch stars. *Astrophys. J.* **684**, 1143–1158 (2008).
20. Bensby, T., Feltzing, S., Oey, M. S. Exploring the Milky Way stellar disk. A detailed elemental abundance study of 714 F and G dwarf stars in the solar neighbourhood. *Astron. Astrophys.* **562**, A71 (2014).
21. Magrini, L. *et al.* The Gaia-ESO survey: radial distribution of abundances in the Galactic disc from open clusters and young-field stars. *Astron. Astrophys.* **603**, A2 (2017).
22. Yong, D., Carney, B.W., Friel, E.D. Elemental abundance ratios in stars of the outer galactic disk. IV. A new sample of open clusters. *Astron. J.* **144**, 95 (2012).
23. Feast, M. W., Menzies, J. W., Matsunaga, N. & Whitelock, P. A. Cepheid variables in the flared outer disc of our galaxy. *Nature* **509**, 342–344 (2014).
24. Kovtyukh, V. *et al.* The chemical composition of Galactic beat Cepheids. *Mon. Not. R. Astron. Soc.* **460**, 2077–2086 (2016).
25. Odenkirchen, M., Grebel, E. K., Dehnen, W., Rix, H.-W. & Cudworth, K. M. Kinematic study of the disrupting globular cluster Palomar 5 using VLT spectra. *Astron. J.* **124**, 1497–1510 (2002).
26. Gratton, R., Sneden, C. & Carretta, E. Abundance variations within globular clusters. *Annu. Rev. Astron. Astrophys.* **42**, 385–440 (2004).
27. Tolstoy, E., Hill, V. & Tosi, M. Star-formation histories, abundances, and kinematics of dwarf galaxies in the Local Group. *Annu. Rev. Astron. Astrophys.* **47**, 371–425 (2009).
28. Gómez, F. A. *et al.* Vertical density waves in the Milky Way disc induced by the Sagittarius dwarf galaxy. *Mon. Not. R. Astron. Soc.* **429**, 159–164 (2013).
29. Laporte, C. F. P., Johnston, K. V., Gómez, F. A., Garavito-Camargo, N. & Besla, G. The influence of Sagittarius and the Large Magellanic Cloud on the Milky Way galaxy. Preprint at <https://arxiv.org/abs/1710.02538> (2017).

Acknowledgements We thank I. Georgiev for help with the telluric correction of the stellar spectrum taken with the UVES spectrograph at the VLT. A.M.S. was supported by grants ESP2015-66134-R and ESP2017-82674-R (MINECO). K.V.J.'s contributions were supported by a grant from the National Science Foundation (AST-1614743). L.C. acknowledges support from the Australian Research Council (grants DP150100250 and FT160100402). Parts of this research were conducted by the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), through project number CE170100013. M.B. acknowledges support from Collaborative Research Center SFB 881 (Heidelberg University, subproject A5) of the Deutsche Forschungsgemeinschaft. C.F.P.L. is supported by a Junior Fellow of the Simons Society of Fellows award from the Simons Foundation. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number OCI-1053575. R.S. is supported by a Royal Society University Research Fellowship. We thank S. Majewski and K. Cunha for interesting discussions on the topic, and J. Bovy for help with implementing the disk-flare profile. We thank T. Müller for assistance with the final, production-quality versions of all figures. We thank the people who realized the Keck Telescope and its instruments and those who operate and maintain the Keck Observatory. We thank the indigenous Hawaiian community for their generous hospitality on their sacred mountain.

Author Contributions This project was led by M.B. and initiated by B.S. and K.V.J. The photometric selection of targets was made by B.S., T.S.L. and A.S. The high-resolution Keck spectra were obtained by J.G.C. The VLT proposal to observe the TriAnd star was prepared by B.S. Spectroscopic analysis of the spectra, including stellar parameters and chemical abundances, was carried out by M.B. L.C. measured stellar effective temperatures and A.M.S. carried out the Bayesian analysis of distances. C.F.P.L. provided the *N*-body simulation that describes the interaction of the Sagittarius galaxy with the Galactic disk. R.S. performed the analysis of stellar kinematics. A.M.P.-W. helped with interpretation of the results and comparison with the models. The manuscript was written mainly by M.B. and A.G. All authors contributed to the text and provided comments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to M.B. (bergemann@mpia-hd.mpg.de).

Reviewer Information *Nature* thanks R. Ibata and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

We acquired high-resolution spectra of 14 red giant branch (RGB) stars in the TriAnd and A13 overdensities^{15,17}. The stars are confirmed members of these overdensities on the basis of their photometry, radial velocities and proper motions. We chose the brightest members of both groups in order to achieve the highest possible signal-to-noise ratios within the time available to us at the Keck Observatory. Metallicity was not a selection criterion. The photometric properties of the observed stars are given in Extended Data Table 1. All observed stars are cool red giants on the upper part of the RGB, and their spectra are generally complex and display strong absorption features caused by molecules, circumstellar shells, and mass loss.

Thirteen stars were observed with the HIRES-R spectrograph at the Keck-1 telescope³⁰ using a spectral resolution of 36,000; one star, TriAnd0_1, was observed using the UVES spectrograph (spectral resolution 47,000) at the VLT. The Keck spectra were taken on the night of 22 October 2016 with typical exposure times of 20–30 min through thin cirrus clouds. The UVES spectrum was taken on the night of 4 September 04, 2016 with a one-hour exposure. All Keck spectra cover the full optical region, from 4,800 Å to 8,770 Å, and the UVES spectrum covers the range from 4,800 Å to 6,800 Å. The signal-to-noise ratio of the HIRES spectra exceeds 200 per spectral resolution element near 5,200 Å at the centre of the echelle order. For the UVES spectrum, the average signal-to-noise ratio at the order centre near 5,500 Å is 50, and it increases to 85 at 6,700 Å. We used the MAKEE pipeline, designed by T. Barlow, to reduce HIRES spectra following standard procedures (bias subtraction, flat fielding, sky subtraction, order extraction and wavelength calibration); we used the ESO Reflex³¹ pipeline to reduce the UVES spectrum.

Detailed spectrum synthesis is essential for determining accurate chemical abundances. We used the MARCS³² stellar model atmosphere grid, because it densely covers the parameter space of the stars that we are interested in. Moreover, it accounts for the necessary molecular opacities, such as MgH³³, that plague the spectra of cool RGB stars. The MARCS models are spherically symmetric for low-gravity giants and line blanketing is treated using the opacity sampling method. The models account for the radiation pressure on molecules, although this does not affect stars in the temperature range and evolutionary stage of our sample.

We determined stellar atmospheric parameters using several techniques. We attempted to follow our own standard procedure³⁴ to derive the atmospheric parameters as closely as possible for the program RGB stars. For cool red giants, spectroscopic estimates based on the excitation–ionization equilibrium of iron provide a reliable indication of the gravity and metallicity. We used the infrared flux method^{35,36} to determine effective temperatures (T_{eff}). Optical and infrared magnitudes were taken from APASS and 2MASS photometry and corrected for interstellar reddening³⁷. We determined surface gravities, metallicities and microturbulent velocities by means of the excitation and ionization balance of the Fe I and Fe II lines. Non-local thermodynamic equilibrium (NLTE) corrections for those stars with surface gravities ($\log(g)$) of about 1 dex and metallicities ($[\text{Fe}/\text{H}]$) above -1 dex are very small³⁸, and barely affect the estimates. The derived stellar parameters and their uncertainties are reported in Extended Data Table 2. We derived uncertainties in T_{eff} using the standard approach³⁶. Uncertainties in $\log(g)$ and $[\text{Fe}/\text{H}]$ are 0.15 dex and 0.1 dex, respectively. They represent the total uncertainty of the method, including the systematic and random error components³⁴.

We computed abundances for the chemical elements O, Na, Mg, Ti, Fe, Ba and Eu, using the least-blended spectral lines that were detected in the observed spectra. Line fitting was done using spectral synthesis with the SME code³⁹. For Mg we used two diagnostic lines, at 5,528 Å and 5,711 Å, adopting experimental data for atomic transitions⁴⁰. For Mg, the mean NLTE abundance corrections are -0.10 dex for the 5,711 Å line and 0.02 dex for the 5,528 Å line⁴⁰. The Fe line list contains 123 lines of Fe I and Fe II (ref. 41). Oxygen abundance was derived using the two forbidden [O I] lines at 6,300 Å and 6,363 Å, with oscillator strengths $\log(gf) = -9.717$ dex and -10.185 dex⁴², respectively. In the UVES spectrum, the 6,300 Å line is contaminated by telluric absorption lines, so the spectrum is first corrected using the ESO Molecfit package⁴³. The combined effects of the departures from the assumptions of one-dimensional hydrostatic equilibrium and local thermodynamic equilibrium (LTE) are negligible for the forbidden oxygen lines in our regime of stellar parameter space⁴⁴.

Na abundances are measured using the features at 5,682 Å and 5,688 Å (for six stars), or 6,154 Å and 6,160 Å (seven stars). This is because slightly different settings are used for different Keck observations, and for a given setting one of the Na I doublets falls in the gap between spectral orders. However, all four Na I features are available in the UVES spectrum of TriAnd0_1, and they give consistent abundances. We also used the 6,154 Å and 6,160 Å features to estimate the Na abundance in the Fornax dwarf spheroidal galaxy⁴⁵. The spectrum for one of the Keck targets is shown in Extended Data Fig. 2, with prominent lines in the region around the 6,154 Å Na I line labelled. The NLTE corrections for Na I lines are

about -0.12 dex for the 6,154 Å and 6,160 Å lines, and around -0.14 dex for the 5,682 Å and 5,688 Å lines⁴⁶.

For Ti, we used 23 lines, including 18 lines of Ti I and five lines of Ti II, which are the least blended by molecular transitions (particularly the molecular transition of MgH, which is a major contaminating species at wavelengths below 6,000 Å). We use LTE Ti abundances, because the NLTE Ti model does not give consistent solutions with one-dimensional hydrostatic models⁴⁷.

We determined Ba abundances using the Ba II lines at 5,853 Å, 6,141 Å and 6,496 Å and applying NLTE corrections⁴⁸, which are in the range -0.03 dex to -0.05 dex. We also took into account isotopic shifts and hyperfine splitting^{49–51}. The only Eu II line that can be measured in the spectra is the feature at 6,645 Å, which is affected by isotopic and hyperfine splitting. The main isotopes are ¹⁵¹Eu and ¹⁵³Eu, with solar abundances of 47.8% and 52.2%, respectively. The isotopic shifts and the hyperfine-splitting magnetic dipole and electric quadrupole constants are taken from experimental studies^{52,53}. Average solar-scaled abundance ratios are given in Extended Data Table 3.

We made every effort to check the accuracy of each abundance measurement; we examined all spectral fits by eye. To estimate the uncertainties in the chemical abundances, we followed our standard procedure⁴⁰. The typical measurement error is about 0.15 dex. Individual abundance errors are given in Extended Data Table 3. If the atomic lines of interest were contaminated by blends, we deemed the measurements unreliable and no abundance is listed. Furthermore, we determined solar abundances using the same line list as for our identified stars; our derived NLTE solar abundances are in very good agreement with reference estimates⁵⁴. Our stellar abundances are taken relative to our solar abundances.

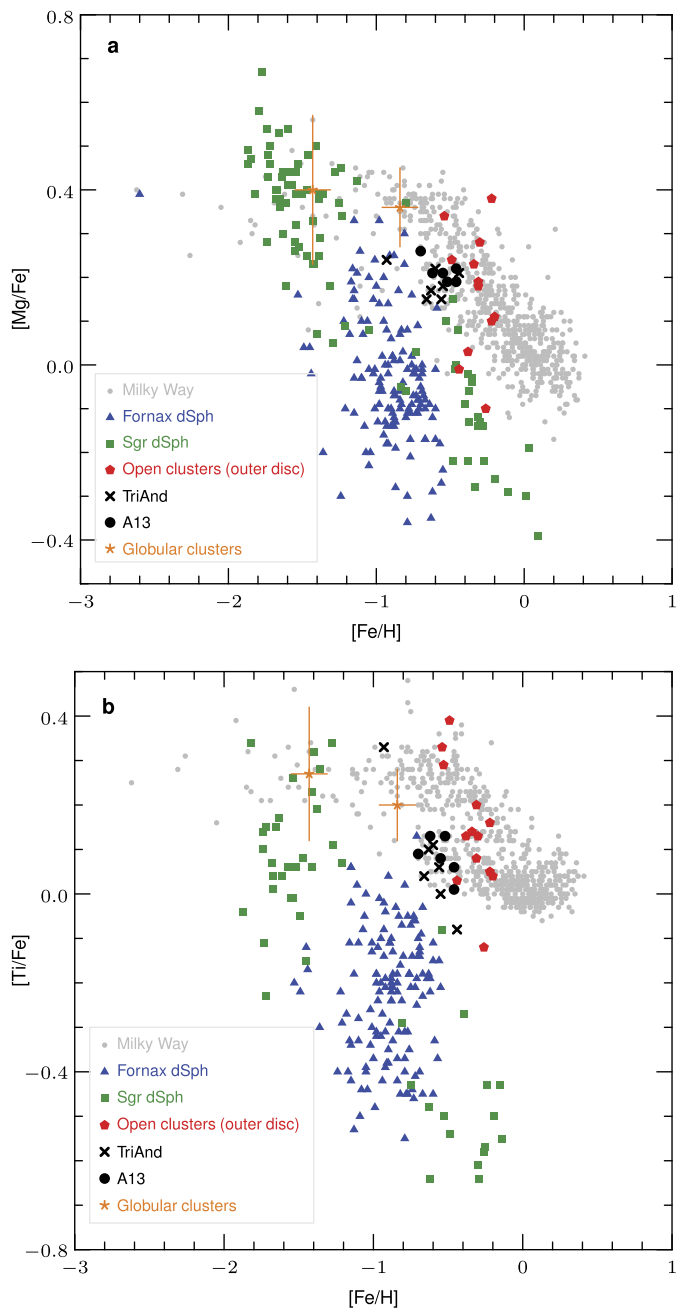
We estimated the distances to the TriAnd and A13 stars by the Bayesian method⁵⁵, using the 2MASS colours and spectroscopic stellar parameters (Fig. 1). The distances are the medians of the posterior probability distribution (PDF) functions, and uncertainties are estimated from the full PDF to a confidence level of 1 s.d. These new distances are about 30% shorter than those determined by our group previously^{15,17}, because that work used an approximate linear relationship between the absolute magnitude and colour of a star, including a parameterized metallicity term. The new distances are not crucial for the present work, because they apply to all stars in the overdensities and do not affect the stellar membership classification. Figure 1 also shows the average disk scale height profile for the low- α stellar population⁵⁶. We derived the line-of-sight velocity dispersion from the measured radial velocities (Extended Data Table 2), after correcting them for the Galactic standard of rest⁵⁷ and for the average motion of stars in the azimuthal direction (Extended Data Fig. 3). The raw data represent the measured radial velocities, from which we first subtract the line-of-sight motion to the Sun—that is, we show the estimated line-of-sight (los) velocity in the Galactic standard of rest (GSR), $V_{\text{GSR}} = V_{\text{los}} + V_{\text{Sun}} \cdot (\mathbf{r} - \mathbf{r}_{\text{Sun}})$, where V_{Sun} is the velocity of the Sun in the Galactic standard of rest⁵⁷, \mathbf{r} is the position vector of the star, and \mathbf{r}_{Sun} is the position vector of the Sun. In a second step, we correct for the projected rotation component of the stellar population: we define $V' = V_{\text{GSR}} + V_{\text{rot}} \cdot (\mathbf{r} - \mathbf{r}_{\text{Sun}})$, where V_{rot} is an estimate for the average motion in the azimuthal direction.

We compare the abundances in Fig. 2 and in Extended Data Fig. 1 with abundances from the following: the Galactic disk and halo stars²⁰; the Sagittarius dwarf spheroidal galaxy^{58–60} (we show only abundances with uncertainties of less than 0.1 dex); the Fornax dwarf spheroidal galaxy^{45,61–63}; globular clusters (M3, ref. 64; and M71, ref. 65); and open clusters Be 25 and NGC 2243 in the outer disk of the Milky Way^{21,22}. $[\text{Fe}/\text{H}]$ and $[\text{Na}/\text{Fe}]$ data for the Galactic disk were derived in NLTE. $[\text{Ba}/\text{Fe}]$, $[\text{Ti}/\text{Fe}]$ and $[\text{Mg}/\text{Fe}]$ data for the Galactic disk stars are LTE estimates. All literature data for dwarf spheroidal systems, globular clusters and open clusters represent LTE estimates. The NLTE corrections tend to reduce $[\text{Ba}/\text{Fe}]$ and $[\text{Na}/\text{Fe}]$ ratios for red giants of sub-solar metallicity. For the dwarf stars, which constitute the comparison sample for the Milky Way disk, the $[\text{Ba}/\text{Fe}]$ NLTE corrections are negligible, within -0.02 dex. The typical NLTE corrections for the Mg I lines in the spectra for dwarf spheroidal galaxies are about 0.15 dex, and for RGB stars are of the order of -0.10 dex (for the 5,711 Å line) or close to zero (for the 5,528 Å line). Therefore, our conclusions would not be affected by the fact that some abundances were derived using LTE.

Data availability. All data relevant to the manuscript are available from the authors. The *N*-body simulation data shown in Fig. 3 are available on request. The data shown in Fig. 1 and in Extended Data Figs 2,3 are included with the paper as Source Data. The data shown in Fig. 2 and in Extended Data Fig. 1 are provided in Extended Data Tables 1–3. HIRES spectra are available at the Keck Observatory Archive, funded by NASA (<https://www2.keck.hawaii.edu/koa/public/koa.php>). UVES data are available from the European Southern Observatory (ESO) Science Archive Facility at http://archive.eso.org/eso/eso_archive_main.html (identification 097.B-0770(A)).

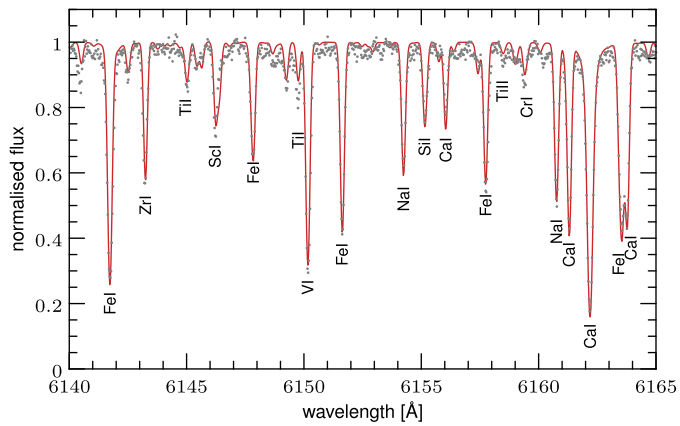
Code availability. The code used to determine stellar parameters and abundances is available at <http://www.stsci.edu/~valenti/sme.html>. The input line list is available on request from the corresponding author. The MARCS model atmospheres are available at the developer's website (<http://marcs.astro.uu.se>). NLTE abundance corrections for the O, Mg, Ti and Fe lines are available at <http://nlte.mpia.de>.

30. Vogt, S. S. *et al.* HIRES: the high-resolution echelle spectrometer on the Keck 10-m telescope. *Proc. SPIE Int. Soc. Opt. Eng.* **2198**, 362 (1994).
31. Freudling, W. *et al.* Automated data reduction workflows for astronomy. The ESO reflex environment. *Astron. Astrophys.* **559**, A96 (2013).
32. Gustafsson, B. *et al.* A grid of MARCS model atmospheres for late-type stars. I. Methods and general properties. *Astron. Astrophys.* **486**, 951–970 (2008).
33. Weck, P. F., Schweitzer, A., Stancil, P. C., Hauschildt, P. H. & Kirby, K. The molecular continuum opacity of ^{24}MgH in cool stellar atmospheres. *Astrophys. J.* **584**, 459–464 (2003).
34. Ruchti, G. R., Bergemann, M., Serenelli, A., Casagrande, L. & Lind, K. Unveiling systematic biases in the 1D LTE excitation-ionization balance of Fe for FGK stars: a novel approach to determination of stellar parameters. *Mon. Not. R. Astron. Soc.* **429**, 126–134 (2013).
35. Casagrande, L., Ramirez, I., Melendez, J., Bessell, M., Asplund, M. An absolutely calibrated Teff scale from the infrared flux method. Dwarfs and subgiants. *Astron. Astrophys.* **512**, A54 (2010).
36. Casagrande, L. *et al.* Towards stellar effective temperatures and diameters at 1 per cent accuracy for future surveys. *Mon. Not. R. Astron. Soc.* **439**, 2060–2073 (2014).
37. Kunder, A. *et al.* The radial velocity experiment (RAVE): fifth data release. *Astron. J.* **153**, 75 (2017).
38. Lind, K., Bergemann, M. & Asplund, M. Non-LTE line formation of Fe in late-type stars—II. 1D spectroscopic stellar parameters. *Mon. Not. R. Astron. Soc.* **427**, 50–60 (2012).
39. Valenti, J. A. & Piskunov, N. E. Spectroscopy made easy. A new tool for fitting observations with synthetic spectra. *Astron. Astrophys. Suppl. Ser.* **118**, 595–603 (1996).
40. Bergemann, M., Collet, R., Amarsi, A. M., Kovalev, M., Ruchti, G., Magic, Z. Non-local thermodynamic equilibrium stellar spectroscopy with 1D and <3D> models. I. Methods and application to magnesium abundances in standard stars. *Astrophys. J.* **847**, 15 (2017).
41. Bergemann, M., Lind, K., Collet, R., Magic, Z. & Asplund, M. Non-LTE line formation of Fe in late-type stars—I. Standard stars with 1D and <3D> model atmospheres. *Mon. Not. R. Astron. Soc.* **427**, 27–49 (2012).
42. Storey, P. J. & Zeippen, C. J. Theoretical values for the [OIII] 5007/4959 line-intensity ratio and homologous cases. *Mon. Not. R. Astron. Soc.* **312**, 813–816 (2000).
43. Smette, A. *et al.* Molecfit: a general tool for telluric absorption correction. I. Method and application to ESO instruments. *Astron. Astrophys.* **576**, A77 (2015).
44. Amarsi, A. M., Asplund, M., Collet, R. & Leenaarts, J. Non-LTE oxygen line formation in 3D hydrodynamic model stellar atmospheres. *Mon. Not. R. Astron. Soc.* **455**, 3735–3751 (2016).
45. Lemasle, B. *et al.* VLT/FLAMES spectroscopy of red giant branch stars in the Fornax dwarf spheroidal galaxy. *Astron. Astrophys.* **572**, A88 (2014).
46. Lind, K., Asplund, M., Barklem, P. S., Belyaev, A. K. Non-LTE calculations for neutral Na in late-type stars using improved atomic data. *Astron. Astrophys.* **528**, A103 (2011).
47. Bergemann, M. Ionization balance of Ti in the photospheres of the Sun and four late-type stars. *Mon. Not. R. Astron. Soc.* **413**, 2184–2198 (2011).
48. Korotin, S. A. *et al.* Grid of theoretical NLTE equivalent widths of four Ba II lines and barium abundance in cool stars. *Astron. Astrophys.* **581**, A70 (2015).
49. Villemoes, P., Arnesen, A., Heijkskjold, F. & Wannstrom, A. Isotope shifts and hyperfine structure of $^{134-138}\text{Ba}$ II by fast ion beam-laser spectroscopy. *J. Phys. B* **26**, 4289–4299 (1993).
50. Van Hove, M., Borghs, G., DeBisschop, P. & Silverans, R. E. J-dependent isotope shifts in the $5d^2D_J$ doublet of barium II. *J. Phys. B* **15**, 1805–1809 (1982).
51. Silverans, R. E., Borghs, G., de Bisschop, P. & van Hove, M. Hyperfine structure of the $5d^2D_J$ states in the alkaline-earth Ba ion by fast-ion-beam laser-rf spectroscopy. *Phys. Rev. A* **33**, 2117–2120 (1986).
52. Lawler, J. E., Wickliffe, M. E., den Hartog, E. A. & Sneden, C. Improved laboratory transition parameters for Eu II and application to the solar europium elemental and isotopic composition. *Astrophys. J.* **563**, 1075–1088 (2001).
53. Villemoes, P., Arnesen, A., Heijkskjold, F., Kastberg, A. & Wannstrom, A. Hyperfine structure measurements of $^{151,153}\text{Eu}$ II with fast ion beam-laser spectroscopy. *Phys. Lett. A* **162**, 178–181 (1992).
54. Asplund, M., Grevesse, N., Sauval, A. J. & Scott, P. The chemical composition of the Sun. *Annu. Rev. Astron. Astrophys.* **47**, 481–522 (2009).
55. Serenelli, A. M., Bergemann, M., Ruchti, G. & Casagrande, L. Bayesian analysis of ages, masses and distances to cool stars with non-LTE spectroscopic parameters. *Mon. Not. R. Astron. Soc.* **429**, 3645–3657 (2013).
56. Bovy, J. *et al.* The stellar population structure of the Galactic disk. *Astrophys. J.* **823**, 30 (2016).
57. Schönrich, R. Galactic rotation and solar motion from stellar kinematics. *Mon. Not. R. Astron. Soc.* **427**, 274–287 (2012).
58. McWilliam, A. & Smecker-Hane, T. A. in *Cosmic Abundances as Records of Stellar Evolution and Nucleosynthesis* (eds Randich, S. & Pasquini, L.) 236–237 (Springer, 2005).
59. Sbordone, L. *et al.* The exotic chemical composition of the Sagittarius dwarf spheroidal galaxy. *Astron. Astrophys.* **465**, 815–824 (2007).
60. Mucciarelli, A. *et al.* Chemical abundances in the nucleus of the Sagittarius dwarf spheroidal galaxy. *Astron. Astrophys.* **605**, A46 (2017).
61. Shetrone, M. *et al.* VLT/UVES abundances in four nearby dwarf spheroidal galaxies. I. Nucleosynthesis and abundance ratios. *Astron. J.* **125**, 684–706 (2003).
62. Kirby, E. N. *et al.* Multi-element abundance measurements from medium-resolution spectra. II. Catalog of stars in Milky Way dwarf satellite galaxies. *Astrophys. J. Suppl. Ser.* **191**, 352–375 (2010).
63. Letarte, B. *et al.* A high-resolution VLT/FLAMES study of individual stars in the centre of the Fornax dwarf spheroidal galaxy. *Astron. Astrophys.* **523**, A17 (2010).
64. Cohen, J. G. & Melendez, J. Abundances in a large sample of stars in M3 and M13. *Astron. J.* **129**, 303–329 (2005).
65. Ramirez, S. V. & Cohen, J. G. Abundances in stars from the red giant branch tip to near the main-sequence turnoff in M71. III. Abundance ratios. *Astron. J.* **123**, 3277–3297 (2002).

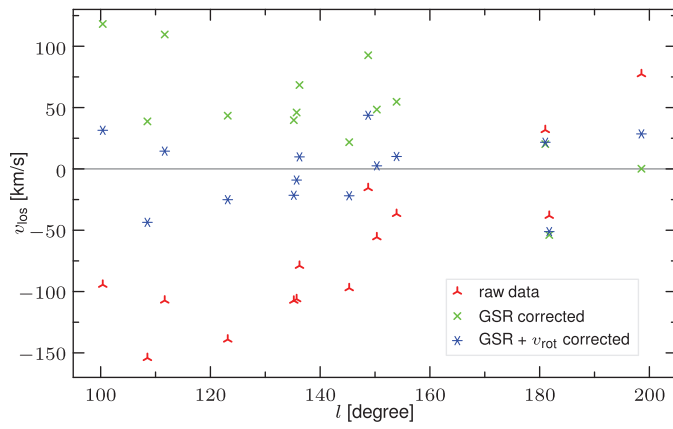


Extended Data Figure 1 | Chemical abundances of the observed stars.

a, b, Chemical abundance ratios $[Mg/Fe]$ (**a**) and $[Ti/Fe]$ (**b**), plotted against metallicity ($[Fe/H]$), in the TriAnd and A13 overdensities, as well as in Milky Way disk and halo stars; the Fornax and Sagittarius (Sgr) dwarf spheroidal galaxies (dSph); open clusters in the Galactic outer disk; and globular clusters (with error bars reflecting intracluster abundance variation, derived as the r.m.s. variance of the sample, with $N = 13$ for the M3 cluster and $N = 25$ for M71). Source data are provided in Extended Data Tables 1 and 3.



Extended Data Figure 2 | Comparison of the observed spectrum of elemental abundances and a model spectrum for a star in the A13 overdensity. Shown are the Keck spectrum of the star 2MASS 07154242+6704006 (grey dots) and the best-fit model spectrum (red line). We used the Na I lines at 6,154 Å and 6,160 Å to determine the Na abundance of the star.



Extended Data Figure 3 | Line-of-sight velocities (v_{los}) of the observed stars, plotted against Galactic longitude (l). GSR, Galactic standard of rest. See Methods for further details.

Extended Data Table 1 | Coordinates and 2MASS photometric magnitudes of the observed stars

2MASS Name	Name	RA	Dec	l	b	J	H	Ks
		(deg)	(deg)	(deg)	(deg)	(mag)	(mag)	(mag)
01552443+4106144	TriAnd5	28.852	41.104	135.70975	−20.175032	11.43	10.58	10.38
02354004+3629079	TriAnd3	38.917	36.486	145.29233	−21.819560	11.38	10.51	10.31
07012973+6526355	A13-20	105.374	65.443	145.29233	25.529054	11.18	10.36	10.18
07031762+6210186	A13-21	105.823	62.172	153.92630	25.133876	11.00	10.21	10.01
07154242+6704006	A13-29	108.927	67.067	148.75062	27.167882	11.09	10.29	10.12
08141720+3952398	A13-05	123.572	39.878	181.02881	32.270140	11.27	10.45	10.29
08182865+2435032	A13-31	124.619	24.584	198.55114	29.322847	11.17	10.34	10.16
07492437+3811176	A13-04	117.352	38.188	181.74952	27.191316	11.21	10.42	10.20
23554397+2901207	TriAnd4	358.933	29.022	108.51723	−32.285203	11.57	10.78	10.57
23484978+4549245	TriAnd7	357.207	45.823	111.67115	−15.674754	10.90	10.01	9.76
00523040+3933030	TriAnd13	13.127	39.551	123.15625	−23.320336	11.64	10.86	10.64
01462028+3604397	TriAnd10	26.585	36.078	135.19653	−25.483016	11.51	10.68	10.47
01540851+3820287	TriAnd2	28.535	38.341	136.23279	−22.905406	11.52	10.66	10.49
23174139+3113043	TriAnd0_1	349.422	31.218	100.37913	−27.515480	11.75	10.92	10.74

RA, right ascension; Dec, declination; l, Galactic longitude; b, Galactic latitude; J, H and Ks, magnitudes in the 2MASS photometric filters.

Extended Data Table 2 | Radial velocities, stellar parameters and distances of the observed stars

2MASS Name	RV	r_{helio}	σ	T_{eff}	$\sigma(T_{\text{eff}})$	$\log(g)$	[Fe/H]	V_{mic}	E(B-V)
	(kms^{-1})	(kpc)	(kpc)	(K)	(K)	(cgs)	(dex)	(kms^{-1})	
01552443+4106144	−106.2	13.73	2.6	3824	140	0.84	−0.63	1.8	0.06
02354004+3629079	−97.2	13.62	2.5	3785	140	0.81	−0.55	1.8	0.05
07012973+6526355	−55.7	9.30	1.8	3927	90	1.16	−0.52	1.7	0.04
07031762+6210186	−36.7	7.59	1.5	4029	117	1.26	−0.55	1.6	0.06
07154242+6704006	−15.6	8.91	1.7	3998	103	1.17	−0.70	1.7	0.04
08141720+3952398	31.7	9.63	1.9	3982	104	1.16	−0.62	1.7	0.04
08182865+2435032	77.3	9.29	1.9	3938	85	1.12	−0.46	1.7	0.03
07492437+3811176	−38.2	9.20	1.8	3930	140	1.16	−0.46	1.7	0.04
23554397+2901207	−154.1	13.54	2.7	3985	133	0.90	−0.63	1.7	0.05
23484978+4549245	−107.3	10.88	1.6	3628	175	0.89	−0.66	1.9	0.12
00523040+3933030	−139.1	13.84	2.6	3888	115	0.93	−0.44	1.8	0.05
01462028+3604397	−107.4	11.40	2.0	3852	117	1.11	−0.56	1.8	0.05
01540851+3820287	−79.0	12.03	2.1	3864	95	1.05	−0.60	1.8	0.05
23174139+3113043	−94.4	14.61	2.3	3909	140	1.02	−0.93	1.7	0.11

RV, radial velocity; r_{helio} , heliocentric distance; cgs, units of surface gravity in the centimetre–gram–second system; E(B-V), extinction coefficient; V_{mic} , microturbulence.

Extended Data Table 3 | Chemical abundances of oxygen, magnesium, sodium, titanium, europium and barium in the observed stars

2MASS Name	[O/Fe]	[Mg/Fe]	[Na/Fe]	[Ti/Fe]	[Eu/Fe]	[Ba/Fe]
01552443+4106144	0.20	0.17	0.11	0.10	0.13	0.19
02354004+3629079	0.17	0.18	0.04	0.00	0.10	0.13
07012973+6526355	0.26	0.19	0.08	0.13	0.18	0.24
07031762+6210186	0.28	0.21	0.10	0.08	–	–
07154242+6704006	0.26	0.26	0.14	0.09	0.28	0.19
08141720+3952398	0.29	0.21	0.13	0.13	0.19	0.12
08182865+2435032	0.18	0.19	0.14	0.06	0.11	0.13
07492437+3811176	0.12	0.22	0.09	0.01	0.12	–0.01
23554397+2901207	0.31	–	0.15	–	0.10	–0.18
23484978+4549245	0.14	0.15	0.05	0.04	0.20	–
00523040+3933030	0.13	0.21	0.08	–0.08	0.10	0.14
01462028+3604397	0.26	0.15	0.11	0.06	0.25	0.27
01540851+3820287	0.26	0.22	0.03	0.11	0.19	0.16
23174139+3113043	0.55	0.24	–0.30	0.33	0.68	0.31

The total (statistical plus systematic) uncertainties in the abundance measurements are: 0.14 dex for Mg, 0.12 dex for Eu, 0.23 dex for Ti, 0.13 dex for Na, 0.18 dex for Ba and 0.21 dex for O (see Methods).

Confocal non-line-of-sight imaging based on the light-cone transform

Matthew O'Toole¹, David B. Lindell¹ & Gordon Wetzstein¹

How to image objects that are hidden from a camera's view is a problem of fundamental importance to many fields of research^{1–20}, with applications in robotic vision, defence, remote sensing, medical imaging and autonomous vehicles. Non-line-of-sight (NLOS) imaging at macroscopic scales has been demonstrated by scanning a visible surface with a pulsed laser and a time-resolved detector^{14–19}. Whereas light detection and ranging (LIDAR) systems use such measurements to recover the shape of visible objects from direct reflections^{21–24}, NLOS imaging reconstructs the shape and albedo of hidden objects from multiply scattered light. Despite recent advances, NLOS imaging has remained impractical owing to the prohibitive memory and processing requirements of existing reconstruction algorithms, and the extremely weak signal of multiply scattered light. Here we show that a confocal scanning procedure can address these challenges by facilitating the derivation of the light-cone transform to solve the NLOS reconstruction problem. This method requires much smaller computational and memory resources than previous reconstruction methods do and images hidden objects at unprecedented resolution. Confocal scanning also provides a sizeable increase in signal and range when imaging retroreflective objects. We quantify the resolution bounds of NLOS imaging, demonstrate its potential for real-time tracking and derive efficient algorithms that incorporate image priors and a physically accurate noise model. Additionally, we describe successful outdoor experiments of NLOS imaging under indirect sunlight.

LIDAR systems use time-resolved sensors to scan the three-dimensional (3D) geometry of objects^{21–24}. Such systems acquire range measurements by recording the time required for light to travel along a direct path from a source to a point on the object and back to a sensor. Recently, these types of sensors have also been used to perform NLOS tracking^{12,13} or imaging^{14–20} of objects 'hidden around corners', where the position and shape of the objects are computed from indirect light paths. The light travelling along indirect paths scatters multiple times before reaching a sensor and may scatter off objects outside a camera's direct line of sight (Fig. 1). Recovering images of hidden objects from indirect light paths involves a challenging inverse problem because there are infinitely many such paths to consider. With applications in remote sensing and machine vision, NLOS imaging could enable capabilities for a variety of imaging systems.

The challenging task of imaging objects that are partially or fully obscured from view has been tackled with approaches based on time-gated imaging², coherence gating³, speckle correlation^{4,5}, wavefront shaping⁶, ghost imaging^{7,8}, structured illumination⁹ and intensity imaging^{10,11}. At macroscopic scales, the most promising NLOS imaging systems rely on time-resolved detectors^{12–20}. However, NLOS imaging with time-resolved systems remains a hard problem for three main reasons. First, the reconstruction step is prohibitively computationally demanding, in terms of both memory requirements and processing cycles. Second, the flux of multiply scattered light is extremely low, requiring either extensive acquisition times in dark environments or a

sufficiently high-power laser to overcome the contribution of ambient light. Finally, NLOS imaging often requires a custom hardware system made with expensive components, thus preventing its widespread use.

Confocal NLOS (C-NLOS) imaging aims to overcome these challenges. Whereas previous NLOS acquisition setups exhaustively illuminate and image pairs of distinct points on a visible surface (such as a wall), the proposed system illuminates and images the same point (Fig. 1) and raster-scans this point across the wall to acquire a 3D transient (that is, time-resolved) image^{14,25–27}. C-NLOS imaging offers several advantages over existing methods. First, it facilitates the derivation of a closed-form solution to the NLOS problem. The proposed NLOS reconstruction procedure is several orders of magnitude faster and more memory-efficient than previous approaches, and it also produces higher-quality reconstructions. Second, whereas indirectly scattered light remains extremely weak for diffuse objects, retroreflective objects (such as road signs, bicycle reflectors and high-visibility safety apparel) considerably increase the indirect signal by reflecting light back to its source with minimal scattering. This retroreflectance property can only be exploited by confocalized systems that simultaneously illuminate and image a common point and may be the enabling factor towards making NLOS imaging practical in certain applications (such as autonomous driving). Third, LIDAR systems already perform confocal scanning to acquire point clouds from direct light paths. Our prototype system was built from the ground up, but commercial LIDAR systems may be capable of supporting the algorithms developed here with minimal hardware modifications.

Similarly to other NLOS imaging approaches, our image formation model makes the following assumptions: there is only single scattering behind the wall (that is, no inter-reflections in the hidden part of the scene), light scatters isotropically (that is, the model ignores Lambert's cosine terms), and no occlusions occur within the hidden scene. Our approach also supports retroreflective materials through a minor modification of the image formation model.

C-NLOS measurements consist of a two-dimensional set of temporal histograms, acquired by confocally scanning points x' , y' on a planar wall at position $z' = 0$. This 3D volume of measurements, τ , is given by

$$\tau(x', y', t) = \iiint_{\Omega} \frac{1}{r^4} \rho(x, y, z) \delta(2\sqrt{(x' - x)^2 + (y' - y)^2 + z^2} - tc) dx dy dz \quad (1)$$

where c is the speed of light. Every measurement sample $\tau(x', y', t)$ captures the photon flux at point (x', y') and time t relative to an incident pulse scattered by the same point at time $t = 0$. Here, the function ρ is the albedo of the hidden scene at each point (x, y, z) with $z > 0$ in the 3D half-space Ω . The Dirac delta function δ represents the surface of a spatio-temporal four-dimensional hypercone given by $x^2 + y^2 + z^2 - (tc/2)^2 = 0$, which models light propagation from the wall to the object and back to the wall. It is also closely related to Minkowski's light cone²⁸, which is a geometric

¹Department of Electrical Engineering, Stanford University, Stanford, California 94305 USA.

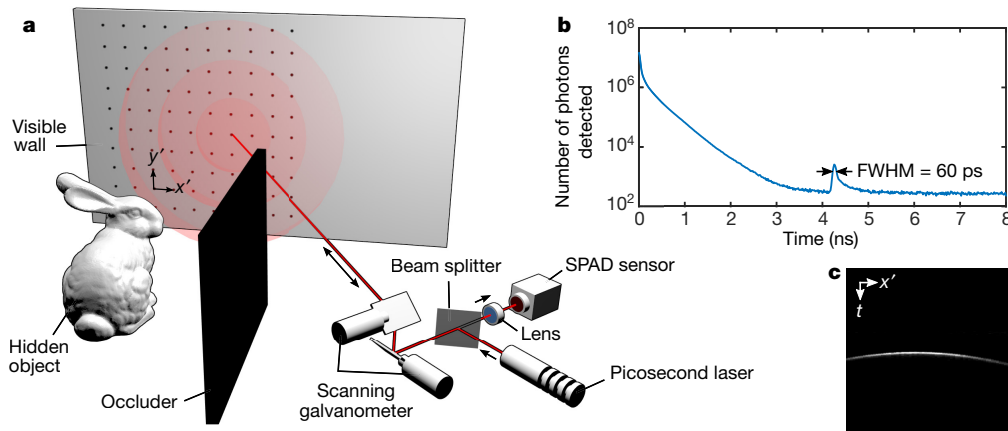


Figure 1 | Overview of confocal imaging hardware and measurements. **a**, A pulsed laser and time-resolved detector raster-scan a wall to record both the direct light reflecting off the wall and the indirect light from a hidden object. **b**, A histogram measured at a scanned point on the visible wall indicates the temporal precision of the detector. In this experiment, the hidden object is a 5 cm × 5 cm square made from retroreflective tape. The detection time of the indirect signal ($t = 4.27$ ns) relative to the direct

signal ($t = 0$ ns) corresponds to twice the distance of the hidden object from the scanned point ($r = 0.64$ m). FWHM, full-width at half-maximum. **c**, Scanning a sequence of points along the wall produces a ‘streak image’ that captures the spatio-temporal geometry of indirect light transport. Each column in this image represents the histogram measured at a discrete point ($x', 0$) on a wall and contains the indirect light from the hidden square.

representation of light propagation through space and time. We note that the function is shift-invariant in the x and y axes, but not in the z axis. A feature of this formulation is that the distance function $r = \sqrt{(x' - x)^2 + (y' - y)^2 + z^2} = tc/2$ can be expressed in terms of the arrival time t ; the radiometric term $1/r^4$ can thus be pulled out of the triple integral. Equation (1) can also be modified to model retro-reflective materials by replacing $1/r^4$ with $1/r^2$, which represents a large increase in the flux of the indirect light (see Supplementary Information for details).

The most remarkable property of equation (1) is the fact that a change of variables in the integral by $z = \sqrt{u}$, $dz/du = 1/(2\sqrt{u})$ and $v = (tc/2)^2$ results in

$$\underbrace{v^{3/2} \tau(x', y', 2\sqrt{v}/c)}_{\mathcal{R}_t\{\tau\}(x', y', v)} = \underbrace{\iint_{\Omega} \frac{1}{2\sqrt{u}} \rho(x, y, \sqrt{u})}_{\mathcal{R}_z\{\rho\}(x, y, u)} \underbrace{\delta((x' - x)^2 + (y' - y)^2 + u - v)}_{h(x' - x, y' - y, v - u)} dx dy du \quad (2)$$

which can be expressed as a straightforward 3D convolution, where $\mathcal{R}_t\{\tau\} = h * \mathcal{R}_z\{\rho\}$. Here, the function h is a shift-invariant 3D convolution kernel, the transform \mathcal{R}_z nonuniformly resamples and attenuates the elements of volume ρ along the z axis, and the transform \mathcal{R}_t nonuniformly resamples and attenuates the measurements τ along the time axis. The inverses of both \mathcal{R}_z and \mathcal{R}_t also have closed-form expressions. We refer to equation (2) as the light-cone transform (LCT).

The image formation model can be discretized as $R_t \tau = H R_z \rho$, where $\tau \in \mathbb{R}_+^{n_x n_y n_t}$ is the vectorized representation of the measurements, and $\rho \in \mathbb{R}_+^{n_x n_y n_z}$ is the vectorized volume of the albedos of the hidden

surface. The process of discretizing each function involves defining a finite grid and integrating the function over each cell in the grid. The matrix $H \in \mathbb{R}_+^{n_x n_y n_h \times n_x n_y n_h}$ represents the shift-invariant 3D convolution operation, and the matrices $R_t \in \mathbb{R}_+^{n_x n_y n_h \times n_x n_y n_t}$ and $R_z \in \mathbb{R}_+^{n_x n_y n_h \times n_x n_y n_z}$ represent the transformation operations applied to the temporal and spatial dimensions, respectively. We note that both transformation matrices are independently applied to their respective dimension and can therefore be applied to large-scale datasets in a computationally and memory-efficient way. Similarly, the 3D convolution operation H can be computed efficiently in the Fourier domain. Together, these matrices represent the discrete LCT.

By treating NLOS imaging as a spatially invariant 3D deconvolution problem, a closed-form solution can be derived from the convolution theorem. The convolution operation is expressed as an element-wise multiplication in the Fourier domain and inverted according to

$$\rho_* = R_z^{-1} F^{-1} \left(\frac{1}{\widehat{H}} \frac{|\widehat{H}|^2}{|\widehat{H}|^2 + \frac{1}{\alpha}} \right) F R_t \tau \quad (3)$$

where F is the 3D discrete Fourier transform, ρ_* is the estimated volume of the albedos of the hidden surface, \widehat{H} is a diagonal matrix containing the Fourier coefficients of the 3D convolution kernel, and α represents the frequency-dependent signal-to-noise ratio of the measurements. This approach is based on Wiener filtering²⁹, which minimizes the mean squared error between the reconstructed volume and the ground truth. As α approaches infinity, the formulation above becomes an inverse filter (that is, the filter applied in the frequency domain is $1/\widehat{H}$). Similarly, the Fourier-domain filter in equation (3) could be replaced by \widehat{H}^* to implement a backprojection reconstruction procedure.

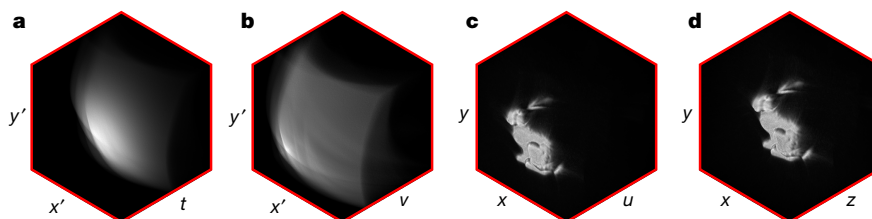


Figure 2 | Overview of the reconstruction procedure. The confocal measurements of the wall **(a)** are resampled and attenuated along the time axis, yielding $R_t \tau$ **(b)**. These measurements are then convolved with

a Wiener filter to produce the volume $R_z \rho_*$ **(c)**, and the result is resampled and attenuated along the depth dimension to produce the hidden volume ρ_* **(d)**. Bunny model from the Stanford Computer Graphics Laboratory.

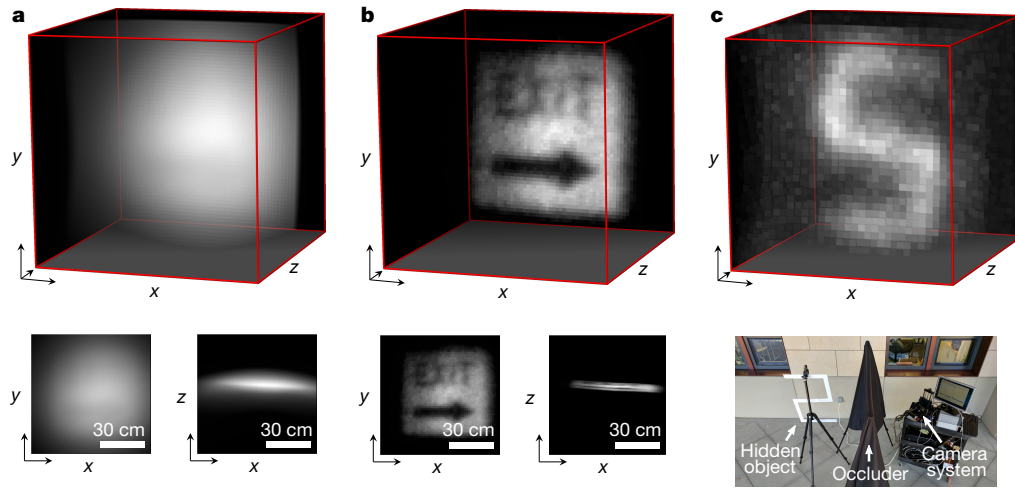


Figure 3 | NLOS reconstructions from SPAD measurements. **a**, Result for a hidden ‘Exit’ sign, obtained using the backprojection method. **b**, Result of the proposed LCT reconstruction procedure. **c**, The proposed method can also reconstruct the shape and albedo of objects outdoors,

under indirect sunlight. The bottom right panel is a photograph of the experimental setup, which consists of a hidden ‘S’-shaped object, black cloth acting as an occluder and the confocal scanning prototype.

Wiener filtering with a constant α inaccurately assumes that the transformed measurements contain white noise. Therefore, we also derive an iterative reconstruction procedure that combines the LCT with a physically accurate Poisson noise model (Supplementary Information).

Figure 2 illustrates the inverse LCT applied to indirect measurements of a bunny model simulated with a physically based ray tracer³⁰. The process involves evaluating equation (3) in three steps: (i) resampling and attenuating the measurements τ with the transform R_b , (ii) applying the Wiener filter to the result, and (iii) applying the inverse transform R_z^{-1} to recover ρ . These three steps are efficient in terms of memory and number of operations required. The most costly step is the application of the Wiener filter, which requires $O(N^3 \log N)$ operations for the 3D fast-Fourier transforms and has memory requirements of $O(N^3)$, where N is the maximum number of elements across all dimensions in space-time. In comparison, existing backprojection-type reconstructions^{15–17} require $O(N^5)$ operations, and methods based on inversion are much more costly both in their memory and processing requirements^{17,18,20}.

In addition to improved runtime and memory efficiency, a primary benefit of the LCT over backprojection-based approaches is that the inverted solution is accurate. In Fig. 3, we compare the reconstruction quality of the backprojection algorithm and the LCT for a retroreflective traffic sign. The dimensions of the hidden sign are $0.61 \text{ m} \times 0.61 \text{ m}$ and the diffuse wall is sampled at 64×64 locations over a $0.8 \text{ m} \times 0.8 \text{ m}$ region. The total exposure time is 6.8 min (that is, 0.1 s per sample) and the runtime for MATLAB to recover a volume of $64 \times 64 \times 512$ voxels is 1 s on a MacBook Pro (3.1-GHz Intel Core i7). To compare the reconstruction quality of the two methods, we compute the backprojection result using the LCT, which is just as

efficient as inverting the problem with the LCT. Even though unfiltered backprojection could be slightly sharpened by linear filters, such as a Laplacian¹⁵, backprojection methods do not solve the inverse problem (see Supplementary Information for detailed comparisons). In Supplementary Information, we also show a variety of reconstructed example scenes, as well as results for NLOS tracking^{11–13} of retroreflective objects in real time.

Applying NLOS imaging outdoors requires the indirect light from the hidden object to be detected in the presence of strong ambient illumination. To accomplish this, C-NLOS imaging takes advantage of the high light throughput associated with retroreflective objects. Figure 3 presents an outdoor NLOS experiment under indirect sunlight (approximately 100 lx). The dimensions of the hidden retroreflective object are $0.76 \text{ m} \times 0.51 \text{ m}$, with 32×32 sampled locations over a $1 \text{ m} \times 1 \text{ m}$ area. The exposure is 0.1 s per sample, with a total exposure time of 1.7 min. MATLAB reconstructs a volume of $32 \times 32 \times 1,024$ voxels in 0.5 s.

The fundamental bounds on the resolution of NLOS imaging approaches couple the full-width at half-maximum of the temporal resolution of the imaging system, represented by the scalar γ , to the smallest resolvable axial Δz and lateral Δx spatial feature size as follows

$$\Delta z \geq \frac{c\gamma}{2} \quad \text{and} \quad \Delta x \geq \frac{c\sqrt{w^2 + z^2}}{2w}\gamma \quad (4)$$

where $2w$ is the sampled width or height of the visible wall (see Supplementary Information for details).

To evaluate the limits of the reconstruction procedure, we simulate the acquisition of $1,024 \times 1,024$ points sampled over a $1 \text{ m} \times 1 \text{ m}$ area

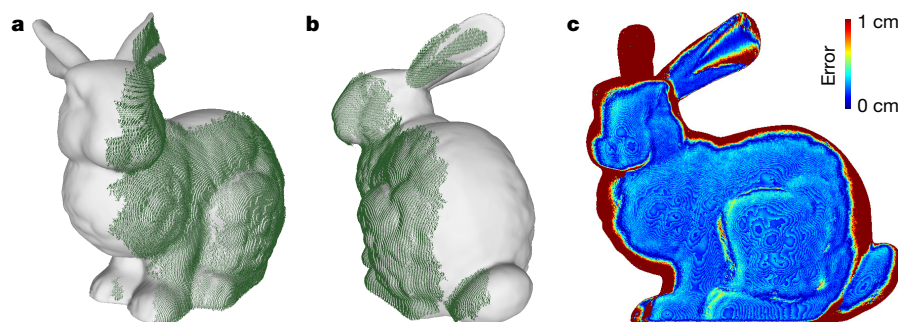


Figure 4 | Comparison between simulated C-NLOS reconstruction and ground-truth geometry. **a**, **b**, Rendered point clouds reconstructed

with the LCT (green) over the ground-truth geometry (grey). **c**, Pointwise difference between the surfaces along the z axis.

and 1,024 time bins with a temporal resolution of 8 ps per bin. We recover a volume containing $1,024 \times 1,024 \times 1,024$ voxels. Figure 4 shows the target geometry in grey and the recovered shape overlaid in green. The error map indicates a median absolute reconstruction error of 2.5 mm (mean absolute error 15.1 mm, mean square error 2.7 mm). Occlusions and higher-order bounces of indirect illumination are not modelled by any existing NLOS imaging method, including ours, which may lead to violations in the image formation model and errors in the reconstructed volume. For example, the right ear of the bunny is not accurately recovered owing to self-occlusions by the left ear in the measurements. We note that the conventional approach of discretizing and inverting the image formation model at this resolution would require an excess of 9 petabytes of memory just to store a sparse representation of the linear system.

The co-design of a confocal scanning technique and a computationally efficient inverse method facilitates fast, high-quality reconstructions of hidden objects. To achieve real-time frame rates with C-NLOS imaging, three improvements to our current prototype are required. First, to reduce acquisition time, a more powerful laser is needed. For eye-safe operation, this laser may need to operate in the short-wave infrared regime^{11,12,22}. Second, for retroreflective objects, the measurement of multiple histograms can be performed in parallel, with minimal crosstalk. This property could enable a single-photon avalanche diode (SPAD) array and a diffused laser source to acquire the full C-NLOS image in a single shot. Third, to improve the computation time, our highly parallelizable algorithm could be implemented in a graphics processing unit or a field-programmable gate array.

The proposed technique thus enables NLOS imaging with conventional hardware at much higher speeds, with a smaller memory footprint and lower power consumption, over a longer range, under ambient lighting and at higher resolution than any existing approach of which we are aware.

Data Availability The measured C-NLOS data and the LCT code supporting the findings of this study are available in the Supplementary Information. Additional data and code are available from the corresponding authors upon request.

Received 29 August 2017; accepted 3 January 2018.

Published online 5 March 2018.

- Freund, I. Looking through walls and around corners. *Physica A* **168**, 49–65 (1990).
- Wang, L., Ho, P. P., Liu, C., Zhang, G. & Alfano, R. R. Ballistic 2-D imaging through scattering walls using an ultrafast optical Kerr gate. *Science* **253**, 769–771 (1991).
- Huang, D. *et al.* Optical coherence tomography. *Science* **254**, 1178–1181 (1991).
- Bertolotti, J. *et al.* Non-invasive imaging through opaque scattering layers. *Nature* **491**, 232–234 (2012).
- Katz, O., Heidmann, P., Fink, M. & Gigan, S. Non-invasive single-shot imaging through scattering layers and around corners via speckle correlations. *Nat. Photon.* **8**, 784–790 (2014).
- Katz, O., Small, E. & Silberberg, Y. Looking around corners and through thin turbid layers in real time with scattered incoherent light. *Nat. Photon.* **6**, 549–553 (2012).
- Stekalov, D. V., Sergienko, A. V., Klyshko, D. N. & Shih, Y. H. Observation of two-photon “ghost” interference and diffraction. *Phys. Rev. Lett.* **74**, 3600–3603 (1995).
- Bennink, R. S., Bentley, S. J. & Boyd, R. W. “Two-photon” coincidence imaging with a classical source. *Phys. Rev. Lett.* **89**, 113601 (2002).
- Sen, P. *et al.* Dual photography. *ACM Trans. Graph.* **24**, 745–755 (2005).
- Bouman, K. L. *et al.* In *IEEE 16th Int. Conference on Computer Vision* 2270–2278 (IEEE, 2017); http://openaccess.thecvf.com/content_iccv_2017/html/Bouman_Turning_Corners_Into_ICCV_2017_paper.html.
- Klein, J., Peters, C., Martin, J., Laurenzis, M. & Hullin, M. B. Tracking objects outside the line of sight using 2D intensity images. *Sci. Rep.* **6**, 32491 (2016).
- Chan, S., Warburton, R. E., Garipey, G., Leach, J. & Faccio, D. Non-line-of-sight tracking of people at long range. *Opt. Express* **25**, 10109–10117 (2017).
- Garipey, G., Tonolini, F., Henderson, R., Leach, J. & Faccio, D. Detection and tracking of moving objects hidden from view. *Nat. Photon.* **10**, 23–26 (2016).
- Kirmani, A., Hutchison, T., Davis, J. & Raskar, R. In *IEEE 12th Int. Conference on Computer Vision* 159–166 (IEEE, 2009); <http://ieeexplore.ieee.org/document/5459160/>.
- Velten, A. *et al.* Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nat. Commun.* **3**, 745 (2012).
- Buttafava, M., Zeman, J., Tosi, A., Eliceiri, K. & Velten, A. Non-line-of-sight imaging using a time-gated single photon avalanche diode. *Opt. Express* **23**, 20997–21011 (2015).
- Gupta, O., Willwacher, T., Velten, A., Veeraraghavan, A. & Raskar, R. Reconstruction of hidden 3D shapes using diffuse reflections. *Opt. Express* **20**, 19096–19108 (2012).
- Wu, D. *et al.* In *Computer Vision – ECCV 2012* (eds Fitzgibbon, A., *et al.*) 542–555 (Springer, 2012); https://link.springer.com/chapter/10.1007/978-3-642-33718-5_39.
- Tsai, C.-Y., Kutulakos, K. N., Narasimhan, S. G. & Sankaranarayanan, A. C. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 7216–7224 (IEEE, 2017); http://openaccess.thecvf.com/content_cvpr_2017/html/Tsai_The_Geometry_of_CVPR_2017_paper.html.
- Heide, F., Xiao, L., Heidrich, W. & Hullin, M. B. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 3222–3229 (IEEE, 2014); http://openaccess.thecvf.com/content_cvpr_2014/html/Heide_Diffuse_Mirrors_3D_2014_CVPR_paper.html.
- Schwarz, B. LIDAR: mapping the world in 3D. *Nat. Photon.* **4**, 429–430 (2010).
- McCarthy, A. *et al.* Kilometer-range, high resolution depth imaging via 1560 nm wavelength single-photon detection. *Opt. Express* **21**, 8904–8915 (2013).
- Kirmani, A. *et al.* First-photon imaging. *Science* **343**, 58–61 (2014).
- Shin, D. *et al.* Photon-efficient imaging with a single-photon camera. *Nat. Commun.* **7**, 12046 (2016).
- Abramson, N. Light-in-flight recording by holography. *Opt. Lett.* **3**, 121–123 (1978).
- Velten, A. *et al.* Femto-photography: capturing and visualizing the propagation of light. *ACM Trans. Graph.* **32**, 44 (2013).
- O’Toole, M. *et al.* In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 2289–2297 (IEEE, 2017); http://openaccess.thecvf.com/content_cvpr_2017/html/OToole_Reconstructing_Transient_Images_CVPR_2017_paper.html.
- Minkowski, H. Raum und zeit. *Phys. Z.* **10**, 104–111 (1909).
- Wiener, N. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series* Vol. 7 (MIT Press, 1949).
- Pharr, M., Jakob, W. & Humphreys, G. *Physically Based Rendering: From Theory to Implementation* 3rd edn (Morgan Kaufmann, 2017).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank K. Zang for his expertise and advice on the SPAD sensor. We also thank B. A. Wandell, J. Chang, I. Kauvar, N. Padmanaban for reviewing the manuscript. M.O.T. is supported by the Government of Canada through the Banting Postdoctoral Fellowships programme. D.B.L. is supported by a Stanford Graduate Fellowship in Science and Engineering. G.W. is supported by a National Science Foundation CAREER award (IIS 1553333), a Terman Faculty Fellowship and by the KAUST Office of Sponsored Research through the Visual Computing Center CCF grant.

Author Contributions M.O.T. conceived the method, developed the experimental setup, performed the indoor measurements and implemented the LCT reconstruction procedure. M.O.T. and D.B.L. performed the outdoor measurements. D.B.L. applied the iterative LCT reconstruction procedures shown in Supplementary Information. G.W. supervised all aspects of the project. All authors took part in designing the experiments and writing the paper and Supplementary Information.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to M.O.T. (matthew.otoole@gmail.com) and G.W. (gordon.wetzstein@stanford.edu).

Reviewer Information *Nature* thanks D. Faccio, V. Goyal and M. Laurenzis for their contribution to the peer review of this work.

Observation of a phononic quadrupole topological insulator

Marc Serra-García^{1*}, Valerio Peri^{1*}, Roman Süssstrunk¹, Osama R. Bilal^{1,2}, Tom Larsen³, Luis Guillermo Villanueva³ & Sebastian D. Huber¹

The modern theory of charge polarization in solids^{1,2} is based on a generalization of Berry's phase³. The possibility of the quantization of this phase^{4,5} arising from parallel transport in momentum space is essential to our understanding of systems with topological band structures^{6–10}. Although based on the concept of charge polarization, this same theory can also be used to characterize the Bloch bands of neutral bosonic systems such as photonic¹¹ or phononic crystals^{12,13}. The theory of this quantized polarization has recently been extended from the dipole moment to higher multipole moments¹⁴. In particular, a two-dimensional quantized quadrupole insulator is predicted to have gapped yet topological one-dimensional edge modes, which stabilize zero-dimensional in-gap corner states¹⁴. However, such a state of matter has not previously been observed experimentally. Here we report measurements of a phononic quadrupole topological insulator. We experimentally characterize the bulk, edge and corner physics of a mechanical metamaterial (a material with tailored mechanical properties) and find the predicted gapped edge and in-gap corner states. We corroborate our findings by comparing the mechanical properties of a topologically non-trivial system to samples in other phases that are predicted by the quadrupole theory. These topological corner states are an important stepping stone to the experimental realization of topologically protected wave guides^{12,15} in higher dimensions, and thereby open up a new path for the design of metamaterials^{16,17}.

A non-vanishing dipole moment \mathbf{p} in an insulator does not lead to any charge accumulation in the bulk. Instead, it manifests through uncompensated surface charges and hence induces potentially interesting surface physics (Fig. 1a). The dipole moment \mathbf{p} is expressible through Berry's phase^{1,3}, which in turn can lead to the quantization of the dipole moment^{4,5,18–21}. All observed topological insulators fit into this framework of quantized dipole moments⁴, or mathematical generalizations thereof²¹. For neutral systems, the abstract quantity \mathbf{p} loses its electromagnetic content. However, it can equally well be used to predict band-structure effects such as stable surface modes. Whether higher-order moments, such as the quadrupole, can lead to distinctly new topological phases of matter has remained unclear.

Recently, a theory for a quantized quadrupole insulator was put forward¹⁴, based on its phenomenology: a bulk quadrupole moment in a finite two-dimensional sample gives rise to surface dipole moments on its one-dimensional edges and to uncompensated charges on the zero-dimensional corners (Fig. 1b). The former indicates gapped edge modes and the latter motivates the presence of in-gap corner excitations. This phenomenology also defines the key technological use of such a quadrupole insulator in mechanical or optical metamaterials: the localized corner modes can be used for acoustic or electromagnetic field enhancements in two dimensions²². Moreover, these states serve as a stepping stone towards topologically protected, one-dimensional channels in three dimensions: when appropriately stacked into three

dimensions, the corner modes give rise to chiral one-dimensional modes along edges of the three-dimensional sample^{17,23–26}.

The phenomenology of gapped edges and gapless corners can be formalized mathematically. Nested Wilson loops have been proposed¹⁴ as a way of obtaining a quantized quadrupole moment (see Methods for details): Wilson loop operators depend only on the bulk properties and encode the edge physics via their eigenvalues $\nu_{\pm}(k_{\alpha})$, $\alpha \in \{x, y\}$, which are known as Wannier bands²⁷. If these Wannier bands $\nu_{\pm}(k_{\alpha})$ are gapped, then the eigenvectors of the Wilson loops can be used to

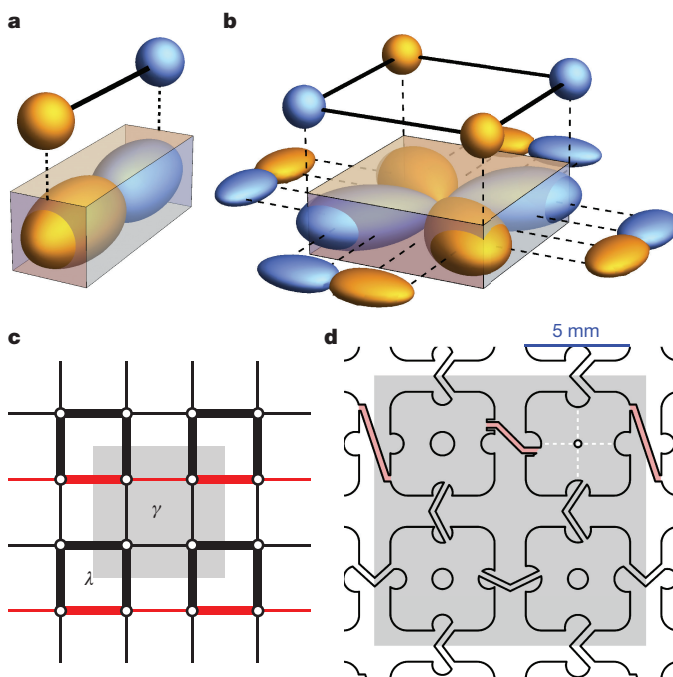


Figure 1 | Quadrupole topological insulator. **a**, In a finite-sized system, a bulk dipole moment induces surface charges as illustrated by the spheres. The colours indicate the charges (positive or negative) of the dipole. The spheres on top illustrate the uncompensated surface charges in a finite system. **b**, A bulk quadrupole moment with its accompanying edge dipoles and corner charges, with the same colour coding as in **a**. **c**, A tight-binding model for a system with a non-vanishing quadrupole moment. Thin (thick) lines denote weak (strong) hoppings with strength γ (λ). The red (black) lines indicate a negative (positive) hopping amplitude. These amplitudes result in a π flux per plaquette. **d**, Metamaterial design that implements the model in **c**. The out-of-plane plate modes with two nodal lines (dashed white lines) are coupled via the bent beams. Beams connecting different sides of a nodal line (shaded red) mediate negative-coupling matrix elements. The grey areas in **c** and **d** mark the unit cell of the tight-binding model.

¹Institute for Theoretical Physics, ETH Zurich, 8093 Zürich, Switzerland. ²Division of Engineering and Applied Science, California Institute of Technology, Pasadena, California 91125, USA.

³Advanced NEMS Group, École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland.

*These authors contributed equally to this work.

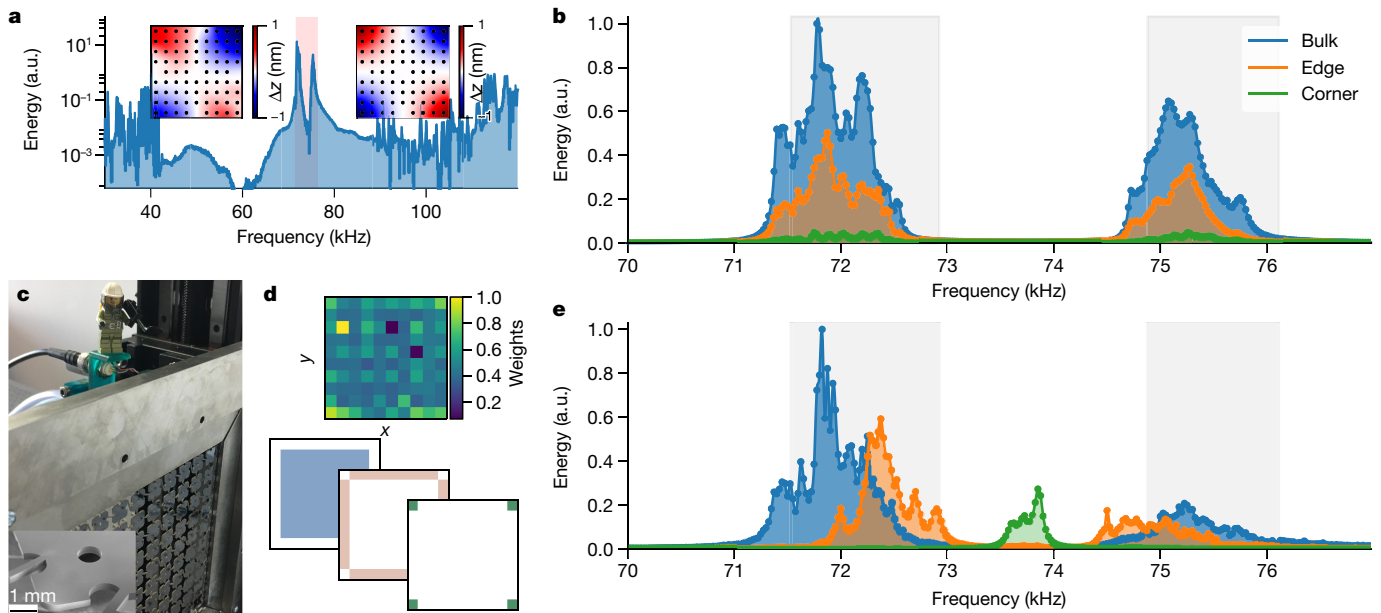


Figure 2 | Quadrupole in-gap states. **a**, Spectrum on a single plate, indicating the large separation between the targeted mode around 73.5 kHz (shaded red) and the bands above and below. The left inset shows the mode profile measured on a single plate (the black dots mark the measurement points used for the interpolation); the right inset shows the numerically calculated mode profile. **d**, The response of all plates at an arbitrary frequency (72.0 kHz). These images are then multiplied by the binary filters shown below to determine the bulk, edge and corner response. Coloured filters are provided for visual identification of bulk,

edge and corner responses in **b** and **e**. **b**, **e**, The resulting spectra for the trivial (**b**) and non-trivial (**e**) sample. For the trivial case (**b**), we see two bands (the grey areas indicate the theoretically predicated locations of the bands) and a central gap with no resonances. For all frequencies, the weights of the bulk (blue), edge (orange) and corner (green) responses are consistent with the fraction of bulk, edge and corner sites in the 10×10 system. The non-trivial case (**e**) shows bulk- and edge-dominated frequency regions and strong corner peaks in the middle of the gap. **c**, Photo of the set-up. The inset shows a close-up of a single plate. a.u., arbitrary units.

define the bulk-induced edge polarization p_{α}^{ν} of the bands below the gap, where α denotes the direction of the polarization. In the same way as for conventional topological insulators⁵, symmetries are required for the quantization of p_{α}^{ν} . In particular, the presence of inversion symmetry I and non-commuting mirror symmetries M_x and M_y leads to a well-defined and quantized $p_{\alpha}^{\nu} \in \{0, 1/2\}$, and the sought-after quantized quadrupole phase is described by¹⁴

$$(p_x^{\nu}, p_y^{\nu}) = \left(\frac{1}{2}, \frac{1}{2}\right) \quad (1)$$

Because a corner terminates two edges, equation (1) could suggest that each of them supports two in-gap states. However, it is an important hallmark of the bulk nature of the quadrupole insulator that each corner hosts only one mode¹⁴ (Fig. 1b).

A tight-binding model for a two-dimensional quantized quadrupole insulator¹⁴ is shown in Fig. 1c. The dimerized hopping with amplitudes λ and γ leads to a bandgap between two pairs of degenerate bands for $\lambda \neq \gamma$ (see Methods). The black (red) lines in Fig. 1c indicate positive (negative) hoppings, effectively emulating a magnetic π flux per plaquette. The π flux requires the mirror symmetry around the horizontal axis (M_y) to be accompanied by a gauge transformation, leading to the non-commutation of M_x and M_y . This model also has inversion I and C_4 rotational symmetry (again up to a gauge transformation), which forces $p_x^{\nu} = p_y^{\nu}$. Moreover, the particle-hole symmetry fixes the corner modes to the middle of the gap. For $\gamma < \lambda$ the topological phase satisfies equation (1), whereas for $\gamma > \lambda$ the trivial phase (0, 0) is realized¹⁴. Here, we seek a mechanical implementation of a quadrupole insulator with $\tilde{x}_i = D_{ij}x_j$, where the dynamical matrix D_{ij} couples local degrees of freedom x_i according to the model in Fig. 1c.

We implement the quadrupole insulator using the concept of perturbative mechanical metamaterials²⁸. The starting point is a single-crystal silicon plate with dimensions $5 \text{ mm} \times 5 \text{ mm} \times 0.364 \text{ mm}$, the mechanical eigenmodes of which are described by the displacement

field $\mathbf{u}(\mathbf{r})$. We work with the first non-rigid-body mode, which is characterized by two perpendicular nodal lines in the out-of-plane component of $\mathbf{u}(\mathbf{r})$ (Figs 1d, 2a). By spectrally separating this mode from the modes below and above it, we can describe the dynamics in some frequency range by specifying only the amplitude x_i of the mode of interest of a given plate i . The hopping elements in D_{ij} are then implemented by thin beams between neighbouring plates. The nodal structure of the mode enables us to mediate couplings of either positive or negative sign, depending on which sides of the nodal lines are connected by the beams. Moreover, the distance to the nodal line controls the coupling strength that is mediated by a given beam. A combinatorial search²⁹ followed by a gradient optimization²⁸ leads to the design in Fig. 1d, which is characterized by a ratio of either $|\gamma/\lambda| = 0.28$ or $|\lambda/\gamma| = 0.28$ (Methods).

All measurements shown are performed using the same scheme. The plates are excited with an ultrasound air transducer. The transducer has a diameter of 5 mm and is in close proximity to the sample, such that only a single plate is excited. We measure the response of the excited plate with a laser interferometer. In this way, we measure the out-of-plane vibration amplitude $\Delta z_i \propto \psi_i^2$, where ψ_i is the eigenmode at the measured frequency (the excitation strength and the measurement both scale with ψ_i). In the insets of Fig. 2a we show the local mode of a single plate measured in this way. In all other figures where energy is shown, we display the mechanical energy $\varepsilon_i \propto \Delta z_i^2$.

To identify the in-gap states we measure $\varepsilon_i(\nu)$ as a function of frequency ν on all plates i . We then apply the filters $\varepsilon_{\alpha}(\nu) = \sum_i \varepsilon_i(\nu) F_{i,\alpha}$ (shown in Fig. 2d) to separate the response of the bulk, the edges and the corners. In Fig. 2b, e we show the resulting spectra for two different samples (Methods). In the topologically trivial case with $\gamma > \lambda$ (Fig. 2b), we observe two frequency bands in which the system absorbs energy (the theoretically predicted location of the bands is indicated in grey). Two features characterize this trivial phase. First, no frequency range is dominated by the edge or corner response. Moreover, the relative weight of the three curves is in accordance with the respective number

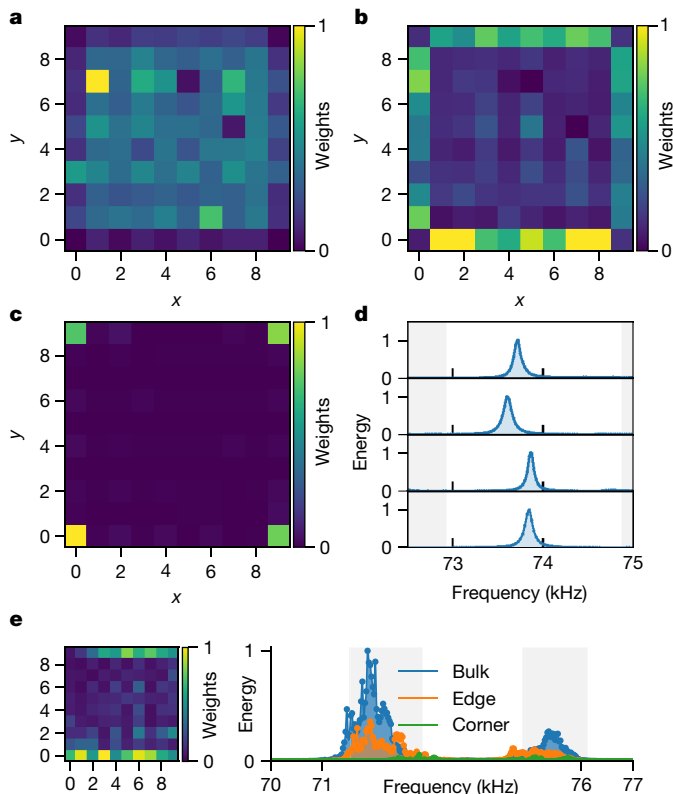


Figure 3 | Edge and corner modes. **a–c**, Normalized integrated weights of the response of the frequency regions in Fig. 1e in which bulk (**a**), edge (**b**) and corner (**c**) modes dominate. **d**, Spectral response of the four corner sites, shown in a clockwise arrangement starting from the top left corner. The combination of gapped edge modes on all four edges (**b**) and the single mode per corner (**d**) evidences the quadrupole nature of our metamaterial. **e**, Spectrum (right) and edge-dominated modes (left) of a system in the non-quadrupole phase $(p_x^-, p_y^-) = (1/2, 0)$, showing no corner states but surface modes on two of the four edges.

of sites in the bulk, edges and corners. Second, no resonances appear in the gap between 72.92 kHz and 74.89 kHz. For the sample with $\gamma < \lambda$ (Fig. 2e), two key-features of the quantized quadrupole phase appear. Close to 72.92 kHz and 74.89 kHz, the response is dominated by the edges, indicative of the bulk-induced gapped edge modes. Sharp resonances at the corners appear in the gap region. A small mirror-symmetry breaking leads to the non-degeneracy of the in-gap states, which we discuss below.

The spectra in Fig. 2b, e enable us to identify three frequency regions \mathcal{B} , \mathcal{E} and \mathcal{C} , in which the bulk (blue), edge (orange) or corner (green) response dominates. To establish the quadrupole nature of the metamaterial, we analyse the site-dependent, frequency-integrated response $\varepsilon_i^\alpha = \sum_{\nu \in \alpha} \varepsilon_i(\nu)$ with $\alpha \in \{\mathcal{B}, \mathcal{E}, \mathcal{C}\}$. In Fig. 3a–c we show the resulting spatial profiles. Note that the bulk induces gapped edge modes on all four sides of the sample.

The hallmark of the quadrupole phase lies in the counting of corner modes: each corner terminates two gapped edges, yet they all host only one in-gap mode¹⁴. In Fig. 3d, we show the response $\varepsilon(\nu)$ for the four corner plates. The resonances in the four corners are split by the presence of next-nearest-neighbour couplings that break the particle-hole symmetry. However, each corner hosts only one resonance peak. Measurements of the Greens functions for the edges further support this claim (Methods).

To corroborate our claim of observing a quadrupole insulator, we explore the phase diagram of ref. 14. When the C_4 symmetry is broken by allowing different hoppings in the x and y directions (Methods), the phase $(p_x^-, p_y^-) = (1/2, 0)$ can be reached via a gap closing of the surface

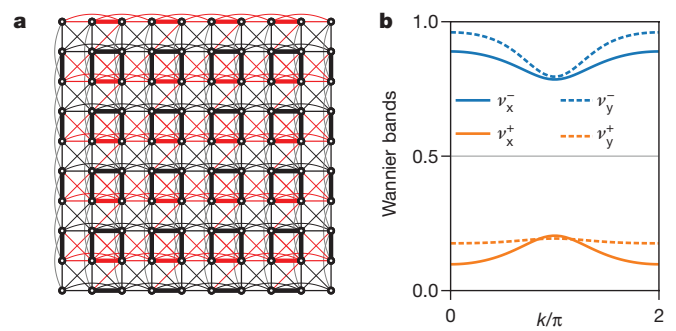


Figure 4 | Reduced model and Wannier bands. **a**, Extracted reduced model for our design. Black (red) lines indicate positive (negative) couplings between the plate modes; the thickness of the lines encodes the hopping amplitude. The unwanted next-nearest-neighbour couplings arise from second-order effects that involve other plate modes, and break the M_x and M_y symmetries. **b**, Calculated Wannier bands from the model in **a**.

modes. The $(1/2, 0)$ phase is characterized by gapped edge spectra on two parallel edges and no emergent edge physics on the perpendicular surfaces¹⁴. Moreover, the induced edge modes are in a trivial state and no corner charges are induced. In Fig. 3e we show measurements on a sample in the $(1/2, 0)$ phase: no in-gap states appear and the frequency region that is dominated by the edges draws its weight from only two surfaces.

In addition to the experimental data presented above, we also validate our system using numerical calculations. The design process for the sample shown in Fig. 1d requires a finite-element simulation of the displacement fields $u_i(\mathbf{r})$ on four unit cells, which contain a total of 16 sites i . The modes obtained in this way can then be projected onto the basis of uncoupled plate modes $u_i^0(\mathbf{r})$. In this way, a reduced-order model \tilde{D}_{ij} in the frequency range of the modes $u_i^0(\mathbf{r})$ is obtained²⁸. In Fig. 4a we show the resulting model extended to a 10×10 system. The nearest-neighbour couplings indeed follow the blueprint of the target model shown in Fig. 1c. However, spurious long-range couplings mediated by off-resonant admixing of other single-plate modes induce a certain amount of mirror-symmetry breaking. This is most notable in the y direction, where negative next-nearest-neighbour couplings are mapped to positive ones, which is not corrected for in the gauge transformation in M_y .

The reduced-order model \tilde{D}_{ij} can also be used to calculate the topological indices (p_x^-, p_y^-) . The gapped Wannier bands $\nu_\pm(k_y)$ and $\nu_\pm(k_x)$ are shown in Fig. 4b. Note that the M_x and M_y symmetries imply that¹⁴ $\nu_+(k_y) + \nu_-(k_y) = 1/2$ and $\nu_+(k_x) + \nu_-(k_x) = 1/2$, respectively. The absence of an exact M_y symmetry indeed leads to a breaking of this rule. This is also reflected in the value of the polarizations: $(p_x^-, p_y^-) = (0.5, 0.56)$. As expected from the structure of \tilde{D}_{ij} shown in Fig. 4a the polarizations are not precisely quantized. However, in-gap corner modes are still observed because the symmetry-breaking terms do not lead to any gap closing, on the edge or in the bulk.

The results presented here underline the power of perturbative metamaterials²⁸. We have used this technique to identify a quantized quadrupole insulator, which represents a new class of topological materials. In addition, a continuous elastic system such as a silicon wafer provides a direct route to technological applications for any theoretical idea that can be represented by a tight-binding model.

Note added in proof: Two preprints^{30,31} that report the observation of topological corner modes have appeared since our initial submission.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 July; accepted 8 November 2017.

Published online 15 January 2018.

1. King-Smith, R. D. & Vanderbilt, D. Theory of polarization of crystalline solids. *Phys. Rev. B* **47**, 1651(R)–1654(R) (1993).
2. Taherinejad, M., Garrity, K. F. & Vanderbilt, D. Wannier center sheets in topological insulators. *Phys. Rev. B* **89**, 115102 (2014).
3. Berry, M. V. Quantal phase factors accompanying adiabatic changes. *Proc. R. Soc. Lond. A* **392**, 45–57 (1984).
4. Thouless, D. J., Kohmoto, M., Nightingale, M. & den Nijs, M. Quantized hall conductance in a two-dimensional periodic potential. *Phys. Rev. Lett.* **49**, 405–408 (1982).
5. Kane, C. L. & Mele, E. J. Z_2 topological order and the quantum spin hall effect. *Phys. Rev. Lett.* **95**, 146802 (2005).
6. Chiu, C. K., Teo, J. C. Y., Schnyder, A. P. & Ryu, S. Classification of topological quantum matter with symmetries. *Rev. Mod. Phys.* **88**, 035005 (2016).
7. Klitzing, K. V., Dorda, G. & Pepper, M. New method for high-accuracy determination of the fine-structure constant based on quantized Hall resistance. *Phys. Rev. Lett.* **45**, 494–497 (1980).
8. König, M. *et al.* Quantum spin hall insulator state in HgTe quantum wells. *Science* **318**, 766–770 (2007).
9. Hsieh, D. *et al.* A topological Dirac insulator in a quantum spin hall phase. *Nature* **452**, 970–974 (2008).
10. Xu, S.-Y. *et al.* Discovery of a Weyl fermion semimetal and topological Fermi arcs. *Science* **349**, 613–617 (2015).
11. Lu, L., Joannopoulos, J. D. & Soljačić, M. Topological states in photonic systems. *Nat. Phys.* **12**, 626–629 (2016).
12. Süsstrunk, R. & Huber, S. D. Observation of phononic helical edge states in a mechanical topological insulator. *Science* **349**, 47–50 (2015).
13. Süsstrunk, R. & Huber, S. D. Classification of topological phonons in linear mechanical metamaterials. *Proc. Natl Acad. Sci. USA* **113**, E4767–E4775 (2016).
14. Benalcazar, W. A., Bernevig, B. A. & Hughes, T. L. Quantized electric multipole insulators. *Science* **357**, 61–66 (2017).
15. Nash, L. M. *et al.* Topological mechanics of gyroscopic metamaterials. *Proc. Natl Acad. Sci. USA* **112**, 14495–14500 (2015).
16. Liu, Z. *et al.* Locally resonant sonic materials. *Science* **289**, 1734–1736 (2000).
17. Huber, S. D. Topological mechanics. *Nat. Phys.* **12**, 621–623 (2016).
18. Bernevig, B. A., Hughes, T. L. & Zhang, S.-C. Quantum spin Hall effect and topological phase transition in HgTe quantum wells. *Science* **314**, 1757–1761 (2006).
19. Fu, L. & Kane, C. L. Time reversal polarization and a Z_2 adiabatic spin pump. *Phys. Rev. B* **74**, 195312 (2006).
20. Moore, J. E. & Balents, L. Topological invariants of time-reversal-invariant band structures. *Phys. Rev. B* **75**, 121306(R) (2007).
21. Qi, X.-L., Hughes, T. L. & Zhang, S.-C. Topological field theory of time-reversal invariant insulators. *Phys. Rev. B* **78**, 195424 (2008).
22. Xiao, M. *et al.* Geometric phase and band inversion in periodic acoustic systems. *Nat. Phys.* **11**, 240–244 (2015).
23. Rechtsman, M. C. *et al.* Photonic Floquet topological insulators. *Nature* **496**, 196–200 (2013).
24. Hafezi, M., Mittal, S., Fan, J., Migdall, A. & Taylor, J. M. Imaging topological edge states in silicon photonics. *Nat. Photon.* **7**, 1001–1005 (2013).
25. Lang, N. & Büchler, H. P. Topological networks for quantum communication between distant qubits. *Quantum Inf.* **3**, 47 (2017).
26. Benalcazar, W. A., Bernevig, B. A. & Hughes, T. L. Electric multipole moments, topological multipole moment pumping, and chiral hinge states in crystalline insulators. *Phys. Rev. B* **96**, 245115 (2017).
27. Fidkowski, L., Jackson, T. S. & Klich, I. Model characterization of gapless edge modes of topological insulators using intermediate Brillouin-zone functions. *Phys. Rev. Lett.* **107**, 036601 (2011).
28. Matlack, K. H., Serra-Garcia, M., Palermo, A., Huber, S. D. & Daraio, C. Designing perturbative metamaterials from discrete models. *Nat. Mater.* <https://doi.org/10.1038/s41563-017-0003-3> (2018).
29. Coullais, C., Teomy, E., de Reus, K., Shokef, Y. & van Hecke, M. Combinatorial design of textured mechanical metamaterials. *Nature* **535**, 529–532 (2016).
30. Imhof, S. *et al.* Topoelectrical circuit realization of topological corner modes. Preprint at <https://arxiv.org/abs/1708.03647> (2017).
31. Peterson, C. W., Benalcazar, W. A., Hughes, T. L. & Bahl, G. Demonstration of a quantized microwave quadrupole insulator with topologically protected corner states. Preprint at <https://arxiv.org/abs/1710.03231> (2017).

Acknowledgements We acknowledge financial support from the Swiss National Science Foundation and the NCCR QSIT. T.L. acknowledges support from a Marie Curie fellowship and O.R.B. from the ETH postdoctoral fellowship FEL-26 15-2.

Author Contributions S.D.H. conceived the research. M.S.-G., V.P. and O.R.B. designed the samples. M.S.-G., V.P., S.D.H. and R.S. conducted the experiments. L.G.V. and T.L. fabricated the samples.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to S.D.H. (sebastian.huber@phys.ethz.ch).

METHODS

Topological quantum number and nested Wilson loops. Here we use the language of fermions, where bands below a given gap can be ‘filled’. For the phononic case, we have to replace ‘filled bands’ with ‘bands below the frequency of interest’. Assuming two bands $n = 1, 2$ are filled, we can use the non-Abelian Berry phase $\mathcal{A}_{nm}^x(\mathbf{k}) = i \langle u_n(\mathbf{k}) | \partial_{k_x} | u_n(\mathbf{k}) \rangle$ of the Bloch wavefunctions $|u_n(\mathbf{k})\rangle$ to construct the Wilson-loop operators

$$\mathcal{W}_x(k_y) = T \exp \left[i \oint \mathcal{A}_{nm}^x(\mathbf{k}) d\mathbf{k}_x \right] \quad (2)$$

Here, T denotes the path ordering along a closed loop in the Brillouin zone. The eigenvalues $\nu^\pm(k_y)$ of $\mathcal{W}_x(k_y)$ are in one-to-one correspondence with the spectrum of an edge perpendicular to the x coordinate²⁷ (or perpendicular to y when x and y are interchanged). If the edge modes are gapped, the eigenvectors $\nu_n^\pm(k_y)$ of $\mathcal{W}_x(k_y)$ can be used to split the filled bands in a well-defined way:

$$|w_\pm(\mathbf{k})\rangle = \sum_{n=1}^2 \nu_n^\pm(k_y) |u_n(\mathbf{k})\rangle$$

The nested polarization is then defined as

$$p_y^{\nu^\pm} = \frac{1}{(2\pi)^2} \int \mathcal{A}_\pm^y(\mathbf{k}) d\mathbf{k}$$

with $\mathcal{A}_\pm^y(\mathbf{k}) = i \langle w_\pm(\mathbf{k}) | \partial_{k_y} | w_\pm(\mathbf{k}) \rangle$. The presence of two mirror symmetries that do not commute (M_x and M_y) is a necessary requirement¹⁴ for the nested polarizations $p_x^{\nu^\pm}$ and $p_y^{\nu^\pm}$ to be quantized to 0 or 1/2.

Model. The model shown in Fig. 1c can be expressed using the Γ matrices¹⁴ $\Gamma_k = -\tau_2 \otimes \sigma_k$ and $\Gamma_4 = -\tau_1 \otimes \sigma_0$, where $k \in \{1, 2, 3\}$ and τ and σ are the standard Pauli matrices:

$$\begin{aligned} D(k_x, k_y) &= [\gamma_x + \lambda_x \cos(k_x)] \Gamma_4 + \lambda_x \sin(k_x) \Gamma_3 \\ &+ [\gamma_y + \lambda_y \cos(k_y)] \Gamma_2 + \lambda_y \sin(k_y) \Gamma_1 \\ &= \sum_{i=1}^4 d_i(\mathbf{k}) \Gamma_i \end{aligned} \quad (3)$$

The C_4 -symmetric version of Fig. 1c is obtained by setting $\lambda_x = \lambda_y$ and $\gamma_x = \gamma_y$. The mirror symmetries are represented by $D(-k_x, k_y) = m_x D(k_x, k_y) m_x^\dagger$ and $D(k_x, -k_y) = m_y D(k_x, k_y) m_y^\dagger$, with $m_x = \tau_1 \otimes \sigma_3$ and $m_y = \tau_1 \otimes \sigma_1$. The eigenvalues of $D(k_x, k_y)$ are given by $\zeta = \pm |\mathbf{d}(\mathbf{k})|$, which leads to two doubly degenerate bands. Bulk gap closings occur when $\mathbf{d}(\mathbf{k}) = \mathbf{0}$, which happens for the C_4 -symmetric case only at $\lambda = \pm \gamma$. The spectrum of the mechanical system is given by $\nu = \sqrt{\nu_0^2 + \zeta}$, with a frequency offset ν_0 . Finally, the eigenvectors $|u_n(\mathbf{k})\rangle$ of $D(k_x, k_y)$ can be used to calculate the Wilson-loop operators of equation (2). The phase diagram and the evolution of the Wannier bands of the model in equation (3) are shown in Extended Data Fig. 1.

The decay of the edge and corner states into the bulk is simple to derive by analogy with the Su–Schrieffer–Heeger model³², in which the wavefunction has a node on every other site and is exponentially decaying with a decay length of $\xi/a = 2/\log(|\lambda/\gamma|)$, where a is the site-to-site distance. For our ratio of $\lambda/\gamma \approx 0.28$ (see below), we obtain $\xi/a \approx 1.6$. In other words, $1 - e^{-4/\xi} \approx 92\%$ of the energy of an edge (corner) mode is stored on the outermost row (corner) site.

Signal analysis. All measurements were performed with an interferometer (IDS3010 from attocube) after exciting with an ultrasound air transducer (SMATR300H19XDA from Steiner & Martins Inc.). All measurements were subject to a systematic uncertainty of the interferometer of about 5 pm, and a statistical error determined by repeated measurements of about 10 pm, resulting in an error estimate on the displacements of around 11.2 pm. Error-propagation analysis results in error bars on all of the data presented in the figures that are smaller than the symbol size. The transducer has an essentially flat frequency response over the frequencies of interest (Extended Data Fig. 2; measured with a second air transducer). The 0.46 dB variations are negligible with respect to the variations in response of 80 dB.

To remove variations in response due to slight misalignments of the measurement point, we normalize the local spectra by

$$\int \Delta z_i(\nu) d\nu \propto \int \psi_i^2(\nu) d\nu$$

as required by the completeness of the eigenmodes. This method is valid only under the assumption that all modes suffer from the same loss or, equivalently, have the same quality factor $Q \approx 1,000$ (determined from the width of the corner modes). This assumption is justified for the following reason. Dissipation arises

from two main sources: the viscoelasticity of the sample and the dissipation into the surrounding air. For both cases, all disconnected plates suffer from the same damping. The perturbative nature of our beams (recall the bandwidth of about 5 kHz around the centre frequency of about 74 kHz), restricts also the effects of the couplings on the dissipation. Our termination is such that all plates see identical surroundings, independent of their location in the bulk, along the edges or on the corners. Moreover, spectra based on data that are not normalized (not shown) are almost identical to those shown in this paper. In all figures where arbitrary units are indicated, we normalize to the maximal value shown in the respective figure.

Because the bulk, edge and corner modes overlap spectrally, there is no unique way to separate them in our measurements. However, because the decay length is extremely short ($\xi/a \approx 1.6$, see above), a separation using the filters shown in Fig. 2d, whereby we simply select sites in the interior, along the edge and at the corner sites, is well justified.

Greens functions. In addition to the measurement of $\psi_i^2(\nu)$ by moving the exciter with the measurement point, we can also measure the Greens function $\psi_i(\nu)\psi_j(\nu)$ by fixing the exciter at site j and moving the measurement point i , while exciting at frequency ν . We first measure the Greens function for the four individual corners at their respective frequencies (determined from Fig. 3d); we show the results in Extended Data Fig. 3. The density maps show the measured wave function $\psi(x, y)$ (x and y replace the site index i). The four panels demonstrate that the four corner modes are independent and that the spread in their frequencies does not arise from their hybridization. Along the edges we show the decay of the wavefunction and compare the envelope of the edges to the theoretical prediction with a decay length of $\xi/a \approx 1.6$.

In Extended Data Fig. 4 we display the analysis of the edge physics by exciting on the bottom left corner and measuring along the lines indicated in Extended Data Fig. 4a. The goal is to show that we can determine the sign of the couplings experimentally. To this end, we model our edge states by using a simple Su–Schrieffer–Heeger model:

$$D(k) = 4\pi^2 \nu_0^2 + \sum_{i=1}^2 d_i(k) \sigma_i$$

where the σ matrices encode the two sublattices, k is the momentum along the edge, and $d_1(k) = \zeta[|\gamma| + |\lambda|\cos(k)]$ and $d_2(k) = \zeta|\lambda|\sin(k)$. Along the horizontal edge, the couplings are positive $\zeta = 1$, whereas along the vertical edge we have negative matrix elements $\zeta = -1$. The spectrum is given by $\omega_\pm(k) = \sqrt{4\pi^2 \nu_0^2 + \zeta |\mathbf{d}(k)|}$, with associated eigenvectors

$$v_\pm(k) = \frac{e^{ikr_i}}{\sqrt{2}} \begin{pmatrix} \pm 1 \\ \frac{d_1(k) + id_2(k)}{|\mathbf{d}(k)|} \end{pmatrix}$$

The highest-frequency modes below the bandgap are given by $\omega_-(\pi)$ for $\zeta > 0$ and $\omega_+(\pi)$ for $\zeta < 0$. For a finite edge, we can build eigenmodes from $v_\pm(k)$ that fulfil the desired boundary conditions. Note that the ± 1 in the first component of $v_\pm(k)$ determines the relative sign between the modes inside one unit cell. Without specifying the exact boundary conditions, or using knowledge of the values of γ and λ , we cannot determine this relative sign. However, we can predict that it will be different on edges with $\zeta = \pm 1$.

To find the frequency of the highest mode per edge below the gap, we show the integrated weight $\bar{\psi}^2 = \sum_i \psi_i^2(\nu)$ along the respective edge in Extended Data Fig. 4b. Fixing the excitation frequencies to the indicated values, we measure the edge wavefunction for these modes. The resulting sign change is indeed different (inside versus between unit cells) on the two edges. Finally, to further justify our filtering, we show that also the decay of the edge modes follows the expected decay with $\xi/a \approx 1.6$.

Sample design. The plate geometries that we investigated were obtained in the framework of perturbative metamaterials²⁸ (Extended Data Fig. 5). We combine geometric elements (silicon plates, beams and holes) to create a material that reproduces the discrete model of ref. 14 over a range of frequencies. A perturbative metamaterial design consists of repeating basic resonating units (5 mm \times 5 mm \times 0.364 mm silicon plates) that weakly interact with neighbouring resonant units. Here, this weak interaction is implemented using thin silicon beams. The weak interaction has two effects: first, the modes of isolated plates hybridize to Bloch bands of small bandwidth, preventing bands originating from unwanted modes to cross in frequency; and second, the weak interaction allows us to approximate the effect of different geometric elements by adding up individual contributions (see ref. 28 for details), resulting in a marked speedup of the calculation times.

The design process starts by establishing a correspondence between the degrees of freedom in the metamaterial and those in the objective discrete model. This is

done by expressing the dynamic deformation of the basic resonant units (plates) of the metamaterial as a linear combination of free-plate eigenmodes (Extended Data Fig. 5b). For sufficiently good spectral separation and sufficiently weak interactions, a single-mode local basis is enough to capture the response of the material with high precision. Each degree of freedom in the objective model is mapped to a single plate, which is assumed to vibrate in its first non-rigid-body mode, which for our parameters has d_{xy} symmetry. Then, we evaluate individual coupling-beam geometries to identify the most suitable designs and create a database relating beam geometry and coupling strength, obtained by simulating two-beam systems (Extended Data Fig. 5c). Geometries are evaluated according to three parameters: (i) the ability to attain a broad range of couplings, (ii) a low compressional strength to prevent the in-plane acoustic bands from reaching high frequencies where they could hybridize with the topological band and (iii) the absence of beam resonances in the frequency range of interest to exclude retardation effects in the couplings. Once the database has been assembled, we start a design by quickly constructing an approximate material geometry, and then refine it by performing a gradient optimization on a full model (Extended Data Fig. 5d) that accounts for the interactions between different geometric features.

We extract the effective theory for our design by first calculating the vibrational eigenmodes of a test system (Extended Data Fig. 5c) using the commercial finite-element method (FEM) package COMSOL Multiphysics. The displacements of the eigenmodes along the three axes u , v and w are then interpolated over a regularly spaced grid with a pitch of 0.05 mm. This interpolation is done for each mode i and plate j , and denoted by ψ_{ijk} . A similar sampling is also performed for individual free-standing plates and denoted φ_k (here, the index k labels the location and component of the displacement that is being interpolated). Once this information has been extracted from finite-element simulations, the displacements of each degree of freedom for each mode are obtained by projecting the test-system displacements into single-plate modes, $\alpha_{ij} = (\varphi_i | \varphi)^{-1} \varphi_k \psi_{ijk}$ (repeated indices denote summation). After this procedure, the components of the matrix α_{ij} contain the displacements of the first non-rigid-body mode of the j th plate for the i th eigenmode of the test system. The use of an interpolated grid enables us to use an individually optimized mesh for each finite-element problem while still being able to express the results of one finite-element simulation in terms of those of another.

The dynamic matrix K that describes the effective theory for the test system is obtained from $K_{lk} = \alpha_{ij} \Omega_{jl}^2 \alpha_{ik}^{-1}$. Here, Ω_{jl}^2 is a diagonal matrix the elements of which contain the square angular frequencies of the modes in the frequency range of interest, $\Omega_{ii}^2 = (2\pi f_i)^2$. The resulting matrix K has the same eigenfrequencies and projected eigenmodes as the full system and therefore provides a good description of the dynamics of the system. This is highlighted in Extended Data Fig. 5g, h, which presents a comparison between the dispersion relation obtained from the effective theory and that obtained by solving a full finite-element model under Bloch boundary conditions.

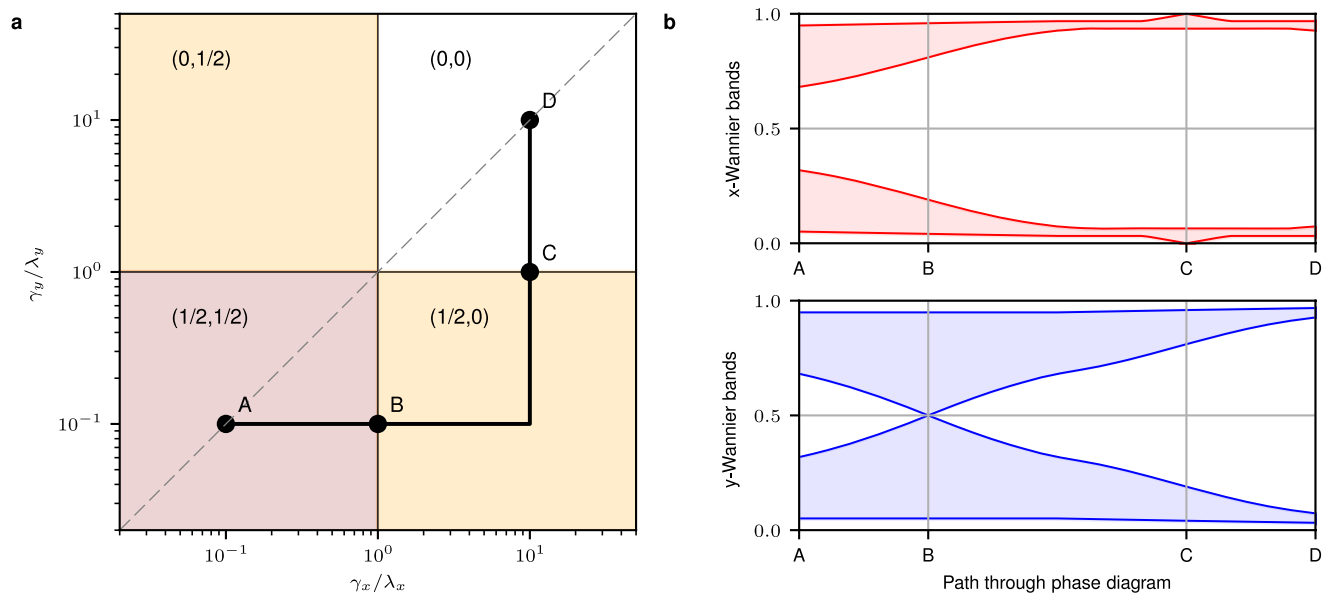
Sample fabrication. The plate and beam geometry of Fig. 1d implements the sought-after weak and strong, positive and negative coupling-matrix elements. The definition of γ as the hopping strength inside a unit cell and of λ as that between unit cells means that $\gamma < \lambda$ is the non-trivial phase. Connected to this identification is the notion of how we are allowed to terminate the system: surfaces have to be compatible with the unit cells; that is, they are not allowed to cut through unit cells. In turn, this means that by using the design in Fig. 1d we can realize all phases shown in this paper by starting from a 10×10 sample in the $(1/2, 1/2)$ phase, then

moving the cut in the y direction by one row of sites to reach the $(1/2, 0)$ phase, and finally moving the termination one column and to end up in the $(0, 0)$ phase. The coupling-matrix elements are given by the ratio of the effective mass density ρ_{eff} of the mode that we use and the stiffness of the beam that connects two plates. We use a 364- μm -thick Si wafer in the (100) orientation, where we align the x and y axes of our model with the in-plane crystalline axes. The mass density of Si is $\rho = 2,330 \text{ kg m}^{-3}$, the Young's moduli are $E_x = E_y = E_z = 130 \text{ GPa}$, the Poisson ratios are $\nu_{yz} = \nu_{zx} = \nu_{xy} = 0.28$ and the shear moduli are $G_{yz} = G_{zx} = G_{xy} = 79.6 \text{ GPa}$ (ref. 33). This results in an offset frequency for our mode of $\nu_0 = 73.895 \pm 0.03 \text{ kHz}$ and coupling-matrix elements of $\lambda = (6.69 \pm 0.17) \times 10^9 (\text{rad s}^{-1})^2$ and $\gamma = (1.89 \pm 0.07) \times 10^9 (\text{rad s}^{-1})^2$. The error estimate is detailed in the next paragraph.

Our samples are fabricated out of double-side-polished 100 mm Si wafers. We measure the thickness of each wafer individually at several points across the wafer and confirm that the overall total thickness variation within each wafer that we use is at most 1 μm . We fabricate plate and beam geometries as illustrated in Fig. 1d using standard micro-fabrication techniques. First, 1 μm of SiO_2 is grown on the wafers via wet thermal oxidation (to be used as an etch mask), and a 2- μm -thick layer of Al (that serves to protect the structure once the whole silicon has been removed) is deposited on the back side of the wafers using e-beam evaporation. A patterned, 5- μm -thick photoresist is used as an etch mask when patterning the front side oxide in a reactive ion etching process. Using the remaining photoresist and the underlying oxide as etch masks, we etch through the wafer with a deep reactive ion etching following a Bosch process, alternating etching and passivation cycles. The ratio between both cycles is chosen to yield vertical side walls. This angle is characterized in several points of each wafer, confirming a variation in the angle of at most 2.5° . The Si etching terminates when the back side oxide is reached. The resulting oxide and aluminium membranes suspended between the beams and plates are removed by wet etching first the aluminium and then the oxide. This step also removes any oxide leftovers present on the front side. The main three sources of error in the targeted model that arise from the sample fabrication are as follows. The first source is the total thickness variation, which we characterize as being less than 1 μm and so represents less than a 0.3% variation across the wafer. The second main source of error comes from different sidewall angles between different parts of the wafer, which we measure to be less than 2.5° . Hence, variation in feature sizes, when comparing front side to back side, may be up to 32 μm . For the width of the plates used, this corresponds to an error of 0.3%. The third source arises from the misalignment of the array with the material crystalline axis (100) . This error has two sources: (a) wafer specifications indicate that the flap is located within $\pm 0.5^\circ$; and (b) specifications of our machine state that the alignment during lithography is around $\pm 1^\circ$. In either case, this results in an overall error of less than 0.1% in the Young's modulus. These errors lead to the stated uncertainties in the local plate frequencies and couplings using standard elasticity theory. Finally, the wafers are clamped between two steel plates (each of 3 mm thickness; Fig. 2c). The impedance mismatch between the steel plates and the wafer leads essentially to fixed boundary conditions $\Delta z = 0$.

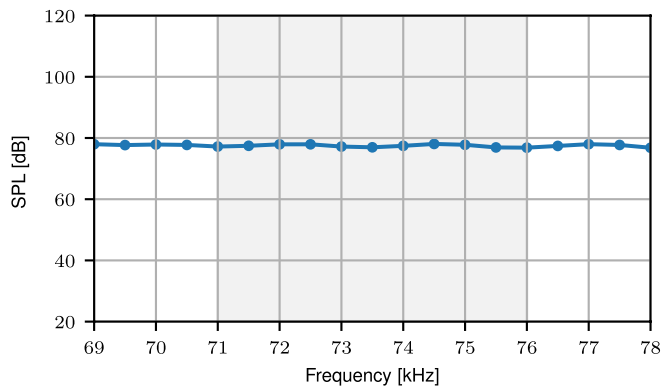
Data availability. The data that support the findings of this study are available from the corresponding author on reasonable request.

32. Su, W. P., Schrieffer, J. R. & Heeger, A. J. Solitons in polyacetylene. *Phys. Rev. Lett.* **42**, 1698–1701 (1979).
33. Hall, J. J. Electronic effects in the elastic constants of n-type silicon. *Phys. Rev.* **161**, 756–761 (1967).

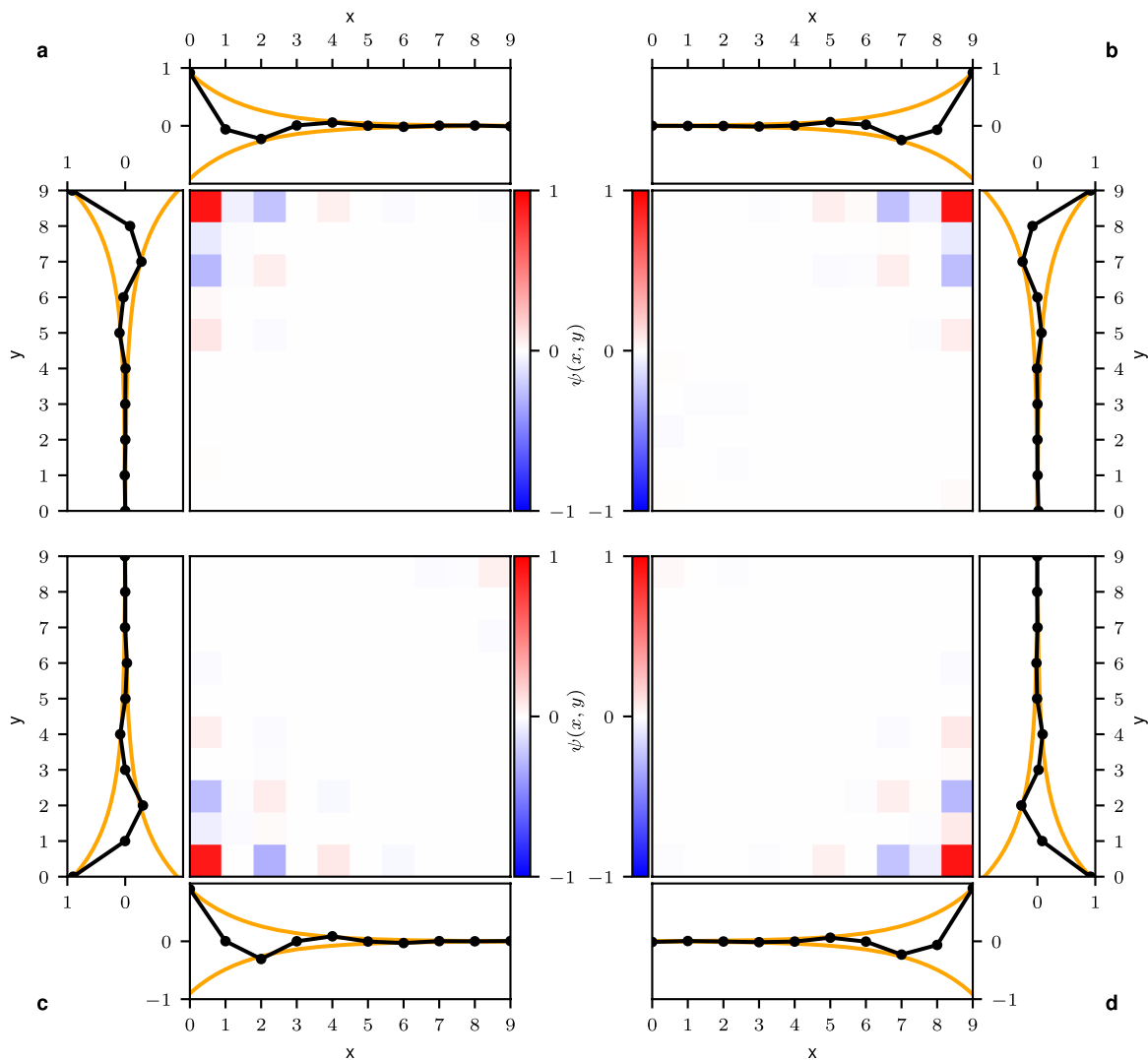


Extended Data Figure 1 | Phase diagram. **a**, Phase diagram of the model in equation (3). The brown area marks the quantized quadrupole phase (1/2, 1/2), whereas the orange areas are the (1/2, 0) and (0, 1/2) phases with no corner modes but emergent edge physics along two parallel edges. The dashed line indicates the C_4 -symmetric line, where the bulk gap is closing at the phase transition. The transitions away from the

C_4 -symmetric line happen through bulk-induced edge transitions, where no bulk gap is closing. **b**, The evolution of the Wannier bands in the x and y directions along the path shown in **a**. The transition from the quadrupole phase to the (1/2, 0) phase is marked by a gap closing at 1/2, removing any polarization in the system. The second transition is induced by a gap closing at 0.

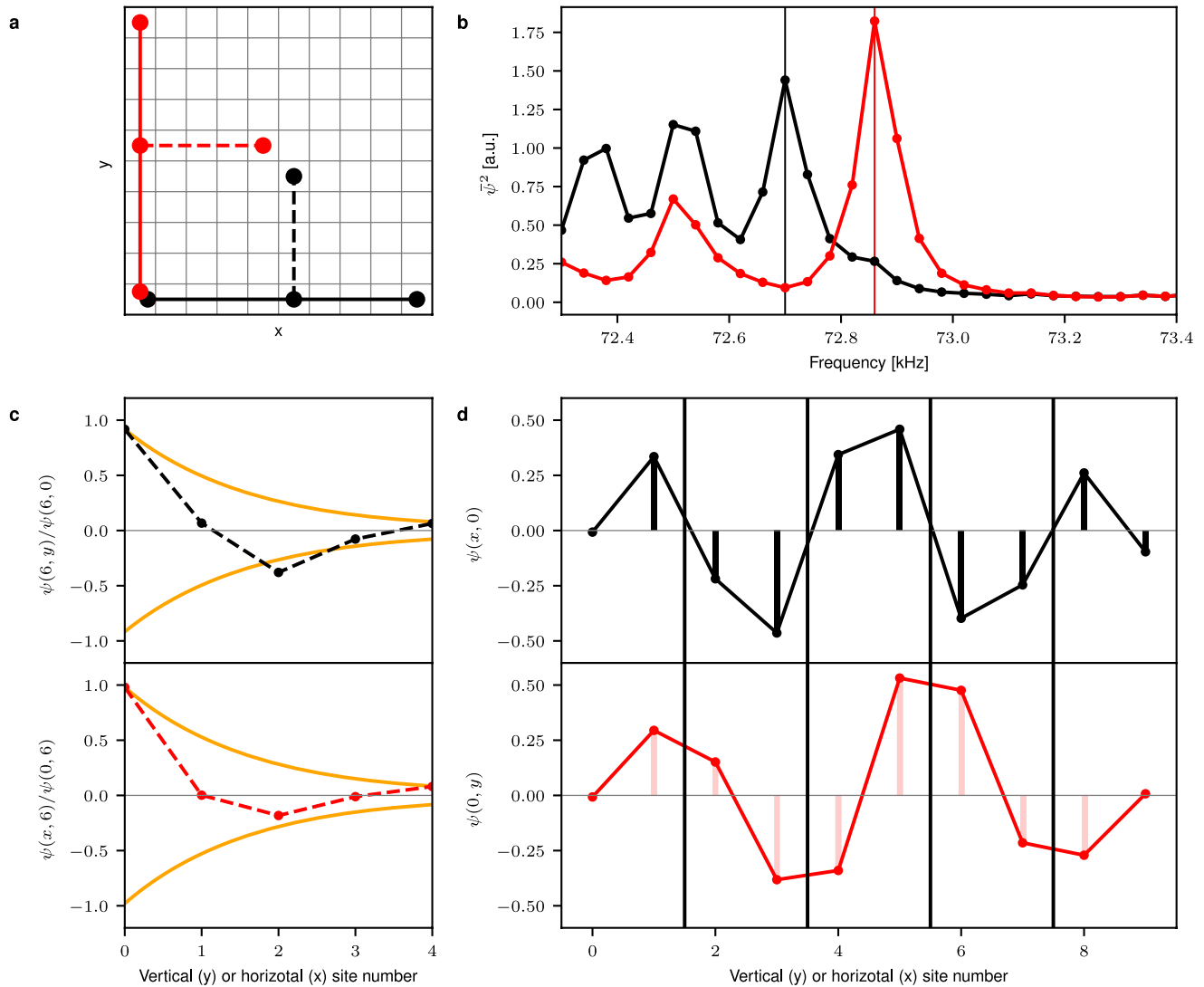


Extended Data Figure 2 | Transducer characterization. The sound pressure level (SPL) in the frequency response of the ultrasound transducer that we used over the frequency region of interest (shaded in grey). The 0.46 dB fluctuations are negligible with respect to the 80 dB variations in the measured response.



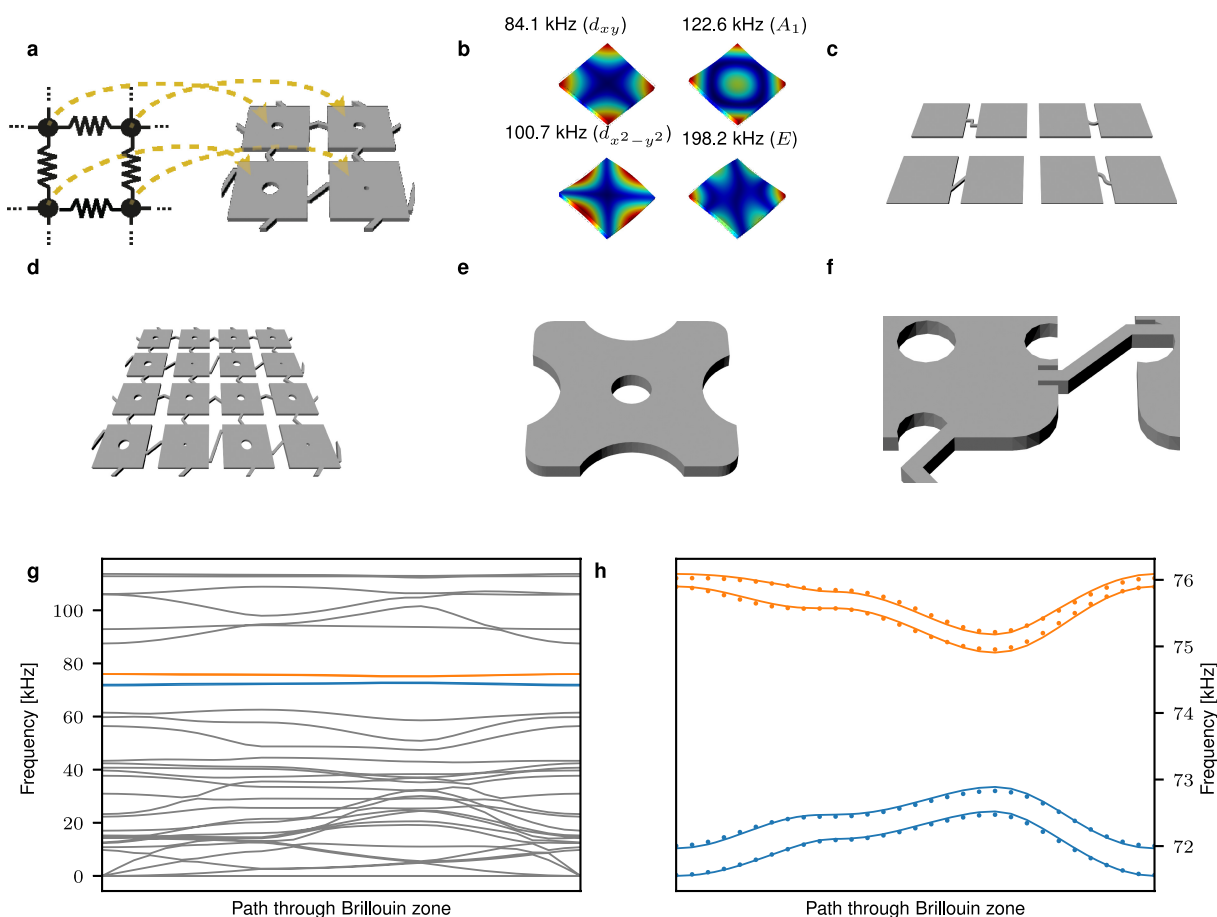
Extended Data Figure 3 | Corner Greens functions. a–d, The corner plates at (0, 9) (a), (9, 9) (b), (0, 0) (c) and (9, 0) (d) are excited at the respective edge-mode frequency. The response recorded (amplitude and phase) enables us to reconstruct the eigenfunctions $\psi(x, y)$ of the individual corner modes. Along each edge, the measured decay of the

modes (black) is shown together with the theoretical prediction (orange). Given the decay length $\xi/a \approx 1.6$, where a is the lattice constant, the residual weight of at most 2% at the corners other than the one that is excited stems from spurious acoustic excitation rather than hybridization.



Extended Data Figure 4 | Edge Greens function. **a**, Sketch of the measured system. The bottom-left plate at $(x, y) = (0, 0)$ was excited. **b**, Integrated frequency response $\bar{\psi}^2 = \sum_i \psi_i^2(\nu)$, where i runs along the red and black edges indicated by the solid lines in **a**. The two highest-frequency peaks below the bandgap (indicated by vertical lines) are analysed. **c**, The decay of the two edge modes into the bulk (black) along

the dashed lines in **a**. The orange lines show the theoretical predictions. **d**, The mode profiles $\psi(x, y)$ along the two edges. The edge with positive couplings has nodes between the unit cells (top panel, black), whereas the edge with negative couplings has nodes inside the unit cell (bottom panel, red), which establishes that a direct measurement of the negative couplings gives rise to the π flux.



Extended Data Figure 5 | Perturbative design of a quadrupole topological insulator.

a, The design approach is based on establishing a correspondence between elements of an objective model (left) and geometric features of the metamaterial (right). Each degree of freedom of the objective model is mapped into a single plate (yellow arrows) by expressing the displacement of each plate as a linear combination of free-plate modes. **b**, Here, only the first non-rigid-body mode, which has d_{xy} symmetry, is used (top left). The other modes are the second, third and fourth non-rigid-body modes. **c**, Independent two-plate systems simulated to create an adequate initial guess for the geometry of the system.

d, Four-unit-cell design simulated during the final gradient optimization. **e**, The refined single-plate design removes material at the maximums of nearby higher-order modes. **f**, Small trenches at the junction between beams and plates. These trenches suppress the coupling to higher-order modes by avoiding regions where these modes have a large displacement. **g**, Dispersion along high-symmetry lines in the Brillouin zone calculated using the finite-element method. The bands that arise from the d_{xy} mode are highlighted in colour. **h**, Detailed view of the spectrum in the frequency range of interest. The dots denote the full finite-element results whereas the lines are calculated from the extracted reduced-order model.

A quantized microwave quadrupole insulator with topologically protected corner states

Christopher W. Peterson¹, Wladimir A. Benalcazar², Taylor L. Hughes² & Gaurav Bahl³

The theory of electric polarization in crystals defines the dipole moment of an insulator in terms of a Berry phase (geometric phase) associated with its electronic ground state^{1,2}. This concept not only solves the long-standing puzzle of how to calculate dipole moments in crystals, but also explains topological band structures in insulators and superconductors, including the quantum anomalous Hall insulator^{3,4} and the quantum spin Hall insulator^{5–7}, as well as quantized adiabatic pumping processes^{8–10}. A recent theoretical study has extended the Berry phase framework to also account for higher electric multipole moments¹¹, revealing the existence of higher-order topological phases that have not previously been observed. Here we demonstrate experimentally a member of this predicted class of materials—a quantized quadrupole topological insulator—produced using a gigahertz-frequency reconfigurable microwave circuit. We confirm the non-trivial topological phase using spectroscopic measurements and by identifying corner states that result from the bulk topology. In addition, we test the critical prediction that these corner states are protected by the topology of the bulk, and are not due to surface artefacts, by deforming the edges of the crystal lattice from the topological to the trivial regime. Our results provide conclusive evidence of a unique form of robustness against disorder and deformation, which is characteristic of higher-order topological insulators.

The simplest model of a system with a quantized dipole moment is a one-dimensional two-band insulator¹². Owing to the presence of chiral or inversion symmetries^{13,14}, this system exhibits quantized fractional edge charges of $\pm e/2$, where e is the electron charge, when its band structure is topological. The fractional edge charges of the quantized dipole insulator are associated with a pair of edge-localized bound states of the Hamiltonian. These edge states have energies that lie within the bulk insulating gap and have been observed in one-dimensional lattices in systems of cold atoms^{15,16} and in several metamaterials^{17–20}. However, the possible existence of quantized higher multipole moments protected by spatial symmetries in crystalline insulators has remained an outstanding question for the past 25 years. A recent theory addressing this issue¹¹ proposes a pair of simple electronic two- and three-dimensional lattice models that exhibit the signatures of quantized electric quadrupole and octupole moments, respectively. A two-dimensional insulator with a quantized quadrupole moment $q_{xy} = e/2$ generates edge-localized dipole moments tangent to the edge of the lattice and corner-localized charges, both of magnitude $e/2$ (Fig. 1a). The corner charges are associated with four corner-localized modes that lie in the middle of the energy gap^{11,21} (Fig. 1c). Although the edge-localized polarizations arise from the gapped, but topological, nature of the edge states, they do not have a spectroscopic manifestation.

Metamaterial analogues of quantum Hall and quantum spin Hall topological insulators have previously been implemented in photonic^{22–24} and phononic^{25,26} systems, as well as in electric circuits²⁷. Here, we implement the two-dimensional quadrupole topological

model from ref. 11. (Fig. 1b) in a metamaterial composed of coupled microwave resonators. Although edge polarizations and corner-localized, topologically protected modes are both signatures of the

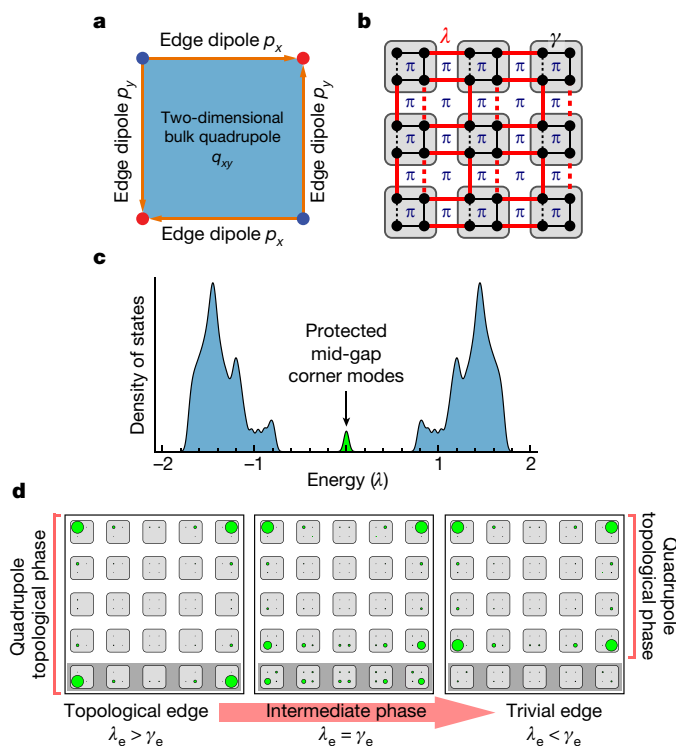


Figure 1 | Quadrupole topological insulator. **a**, Two-dimensional bulk quadrupole topological insulator (blue square) with edge-localized topological dipoles (orange arrows) and corner-localized charges of $\pm e/2$ (red and blue dots). **b**, Tight-binding representation of a quadrupole topological insulator with four sites per unit cell. Red lines denote coupling between unit cells (coupling rate, λ) and black lines represent couplings within unit cells (γ). Dashed lines indicate a $-\pi$ phase factor on the coupling, a gauge choice for the creation of a synthetic magnetic flux of π per plaquette. The insulator is in the quadrupole topological phase for $\lambda > \gamma$ and in the trivial phase for $\lambda < \gamma$. **c**, Theoretically calculated density of states for the quantized quadrupole insulator (5×5 unit cells) shown in **b** with fully open boundaries. The lower and upper bands (blue) have eigenstates delocalized in the bulk, whereas the states in the middle of the gap (green) are confined to the corners, as shown in **d**. The energy is expressed in units of λ . **d**, Theoretically calculated probability density functions (green circles; the areas of the circles correspond to the probability) of the four in-gap modes during reconfiguration of the lowest-edge unit cells from $\lambda_e/\gamma_e = 4.5$ (left) to $\lambda_e/\gamma_e = 1$ (centre) and to $\lambda_e/\gamma_e = 1/4.5$ (right). Throughout the deformation, only γ_e changes, while $\lambda_e = \lambda = 1$ and $\gamma = 1/4.5$. Experimental test results of this deformation are shown in Fig. 4.

¹Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801-2918, USA. ²Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801-2918, USA. ³Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801-2918, USA.

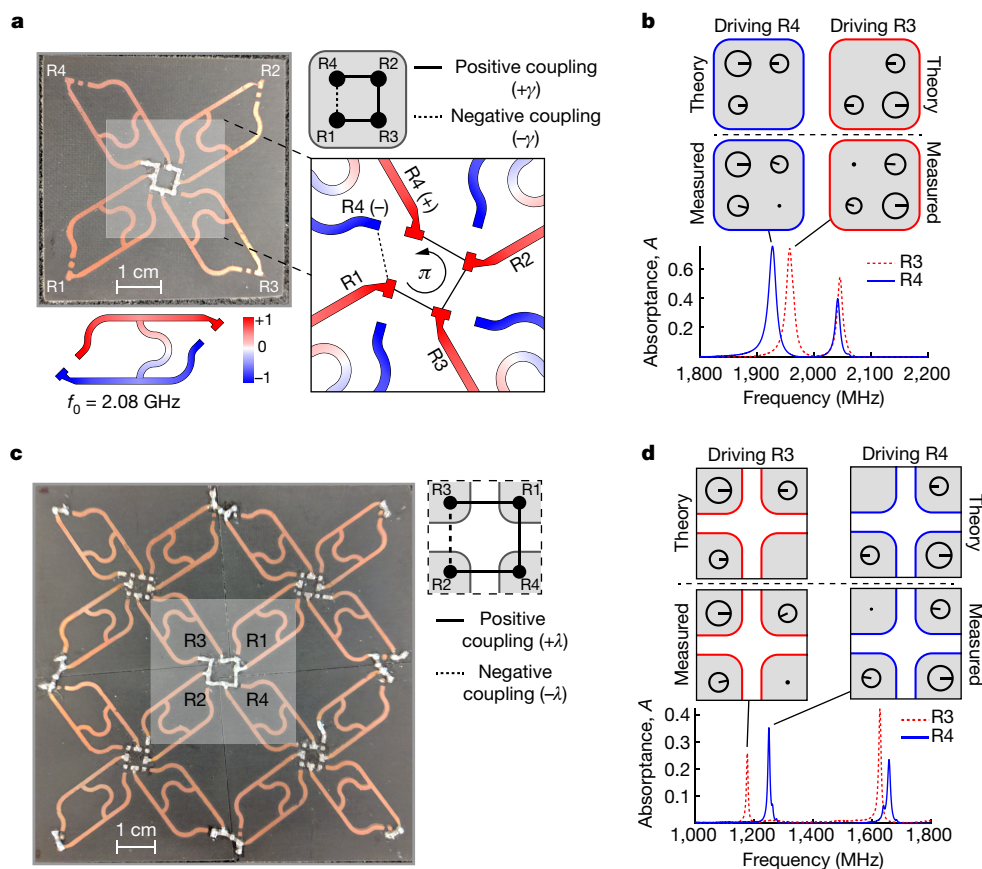


Figure 2 | Verification of microwave quadrupole lattice bulk topology.

a, A unit cell of the quadrupole topological insulator (top left, photograph; top right, schematic) is composed of four H-shaped microstrip resonators that are capacitively coupled. Each resonator has a fundamental mode at $f_0 = 2.08$ GHz (colours represent normalized voltage amplitude; bottom left). The coupling between resonators R4 and R1 adds an extra phase shift of π (negative coupling), as shown in the detailed schematic (bottom right), to produce a π flux through the unit-cell plaquette. $\gamma \approx 35$ MHz. **b**, Eigenmode verification for the unit-cell plaquette. The resonator frequencies are shifted to about 2 GHz owing to capacitive loading from the coupling capacitors. The theoretical and measured eigenmodes are

presented as phasor diagrams: the circle diameter corresponds to the magnitude of the resonator excitation, the line corresponds to the phase (0 is on the right and increases anticlockwise). When driving R4, R1 is in-phase, confirming the negative coupling between R1 and R4. **c**, A 2×2 test array of unit cells with $\gamma \rightarrow 0$. Negative coupling is set between R2 and R3, as illustrated in the schematic on the right. $\lambda \approx 150$ MHz. **d**, The eigenmodes of a plaquette formed by the four central resonators of a 2×2 array are similar to those of the unit cell because of the π flux. When driving R3, resonator R2 is in phase, confirming the negative coupling between these resonators. Here, the resonator frequencies are shifted to about 1.4 GHz owing to the greater capacitive loading than in the case of **b**.

non-trivial topology of the bulk, we focus on the latter, as these modes can provide direct spectroscopic evidence of the existence of the non-trivial quadrupole topological phase. Specifically, we experimentally demonstrate the existence of mid-gap modes that are localized at the corners of the lattice (Fig. 1d, left). Furthermore, we provide evidence that these corner modes are not due to surface effects, but are required by the topological bulk phase. We accomplish this by deforming one of the edges of the lattice from the topological to the trivial regime, and we observe that the mid-gap corner modes are not destroyed; instead, they recede into the sample, towards the corners on the newly generated boundaries of the quadrupole topological phase (Fig. 1d).

The microwave quadrupole topological insulator studied here consists of a square lattice of unit cells, in which each unit cell is composed of four identical resonators (Fig. 1b). The coupling rates γ and λ describe coupling between resonators within the same unit cell and between adjacent unit cells, respectively. Each plaquette, a square of any four adjacent resonators within or between unit cells, contains a single coupling term that carries an extra phase shift of π (dashed lines in Fig. 1b), which amounts to the generation of a synthetic magnetic π flux threading the plaquette (equivalent to half the magnetic flux quantum, $\frac{1}{2}\Phi_0 = h/(2e)$, where h is the Planck constant). The existence of this non-zero flux opens both the bulk and the edge spectral gaps, which are necessary to maintain the corner-localized mid-gap modes.

The resonators in our experimental array are H-shaped microstrip transmission lines that have a fundamental resonance at $f_0 = 2.08$ GHz, a typical linewidth of about 15 MHz and the spatial voltage distribution shown in Fig. 2a (bottom). At the centre of the cross-piece lies a voltage node and the end-points of the resonator are a quarter-wavelength away from the centre—they are therefore anti-nodes. Adjacent tips are separated by a half-wavelength and thus differ in phase by π . The resonators are designed so that anti-nodal points with opposite phases are physically close to each other, which facilitates the coupling of adjacent resonators with either no extra phase shift or an additional phase shift of π (negative coupling). To produce the quadrupole topology, in each plaquette we arrange three couplings as positive and one coupling as negative, as shown in Fig. 2a.

We experimentally confirm that a π flux threads each plaquette by examining the limiting cases of $\lambda \rightarrow 0$ and $\gamma \rightarrow 0$. This experimental verification is necessary to ensure that the spectral features that we measure are due to the bulk quadrupole topology, because corner modes—even topologically protected ones—are not unique to the quadrupole topological insulator^{21,28,29}. In the $\lambda \rightarrow 0$ limit, the array consists of isolated unit-cell plaquettes, as shown in Fig. 2a. The behaviour of the array can be predicted theoretically by a direct diagonalization of the four-site Hamiltonian, a tight-binding representation of which is shown by the grey unit cell in Fig. 2a. For a coupling rate of γ between all resonators and a π flux threading the plaquette, the

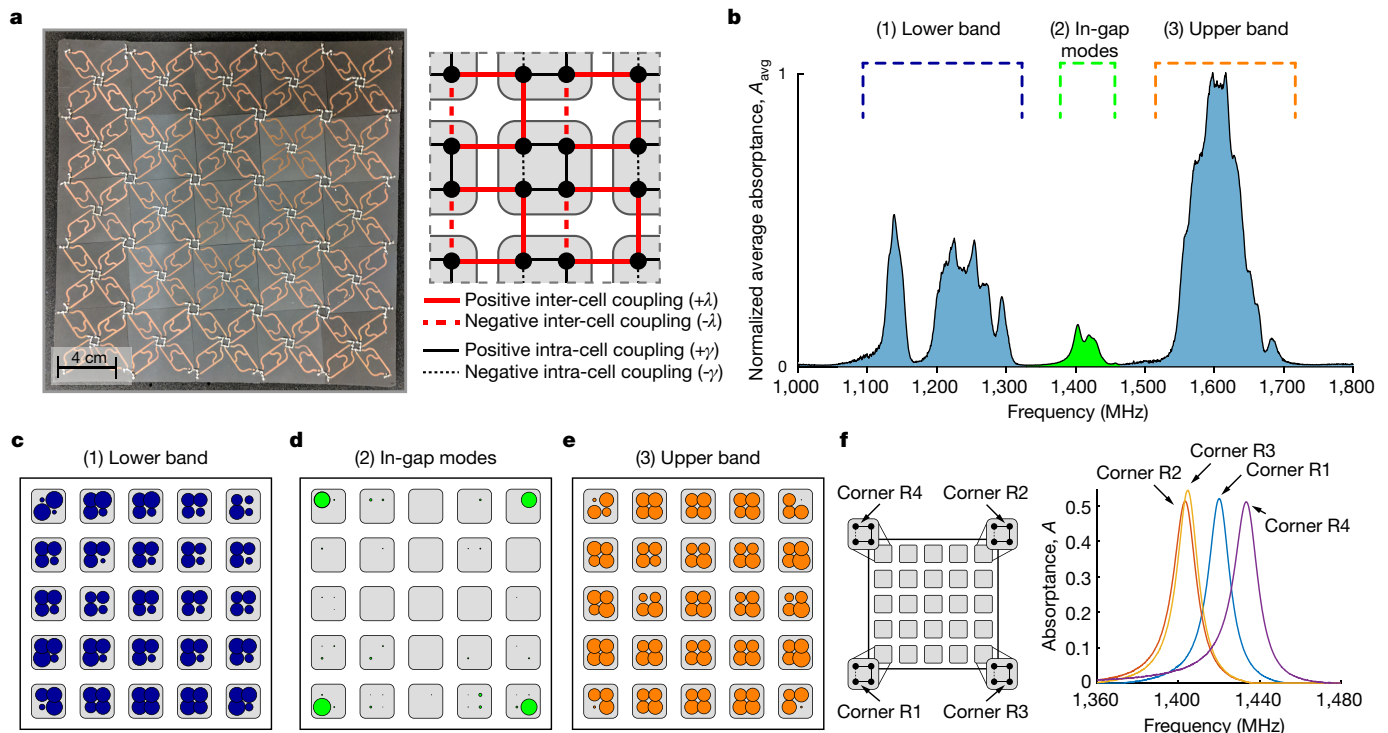


Figure 3 | Demonstration of microwave quadrupole topological insulator. **a**, Photograph of the experimental array of coupled resonators that form a quadrupole lattice. The array has 5×5 unit cells (Fig. 2a). We set the couplings to a ratio $\lambda/\gamma \approx 4.3$. The schematic on the right shows the connectivity of a bulk unit cell. **b**, Normalized average absorbance spectrum of all the resonators in the array (see Methods for details). We observe two large bands (blue) separated by a bandgap containing in-gap modes (green). **c**, Spatial distribution of absorbance summed over the lower frequency band indicated in **b**. Within this band, the response

eigenfrequencies are $\pm\sqrt{2}\gamma$, each of which is two-fold degenerate (see Methods). Because the non-trivial topology of the full array is clearly manifest in either the upper or the lower band (see Fig. 3b), we choose to characterize only the lower band, at $-\sqrt{2}\gamma$.

The measured power absorbance spectrum (the ratio of absorbed power to incident power) of an isolated unit cell is shown in Fig. 2b; see Methods for details on the measurement technique. As predicted by the Hamiltonian diagonalization, we find two pairs of nearly degenerate modes, separated by 88 MHz (measured on resonator R3) and 114 MHz (measured on R4). The discrepancy in mode frequency is due to asymmetric capacitive loading, which was present throughout our experiments (see Methods). The spatial distribution of the lower pair of modes is measured through the voltage amplitude and phase response at each resonator in the plaquette when either R3 or R4 is stimulated (see Methods). We find good agreement between the magnitudes and phases of the theoretical and measured modes (Fig. 2b): their normalized overlap integrals, calculated as an inner product of these modes, are about 0.98 and 0.97 when measured on R3 and R4, respectively. Characteristic mode shapes appear owing to destructive interference (caused by the π flux) between counter-circulating paths around the plaquette. Specifically, when R4 is excited, the mode vanishes for the diagonal resonator R3 (and vice versa). In Methods, we discuss the clear contrast of this situation with the modes predicted for plaquettes with zero flux, although the two cases can exhibit spectral similarities.

In the $\gamma \rightarrow 0$ limit, the array consists of isolated inter-unit-cell plaquettes (Fig. 2c, highlighted region). These plaquettes are nearly identical to the isolated unit cell, but the negative coupling is placed between R2 and R3 and the coupling rate λ is larger. We experimentally verify that the eigenmodes of this inter-unit-cell plaquette also have the features expected for a π flux by performing similar measurements to

is dominated by bulk and edge resonators. The areas of the circles correspond to local absorbance. **d**, Spatial distribution of absorbance summed over the in-gap band noted in **b**. The in-gap modes are localized only on the corner resonators, which are not excited in the lower or upper band. **e**, Spatial distribution of absorbance summed over the upper band indicated in **b**, showing excitation of the bulk and edge resonators. **f**, Individual absorbance spectra of the corner resonators reveal that each corner resonator supports only a single mode.

the single-unit-cell case (Fig. 2d). For this measurement, the capacitors that originally coupled the resonators within the unit cells are removed to ensure that $\gamma = 0$. Although the modes are again not perfectly degenerate, we find good agreement between the theoretical and measured mode shapes: their normalized overlap integrals are about 0.96 and 0.99 when measured on R3 and R4, respectively. The frequency separations of 450 MHz (measured on R3) and 408 MHz (measured on R4) between the two pairs of modes is on average about 4.3 times larger than in the isolated unit cell, revealing that the average ratio of the coupling rates is $\lambda/\gamma \approx 4.3$.

After the experimental verification of the plaquette building blocks, we construct a quadrupole topological insulator using a 5×5 array of unit cells (Fig. 3a) with coupling ratio $\lambda/\gamma \approx 4.3$ and the topology shown in Fig. 1b. The power absorbance spectrum of each resonator in the full array is measured in the same way as in the isolated plaquettes. The average absorbance across the entire array is presented in Fig. 3b. Three spectral bands are identifiable: broad lower and upper bands (blue) separated by a bandgap, and a narrow band of modes near the centre of the bandgap (green). The spatial distributions of these three bands, obtained by summing over each band, are shown in Fig. 3c–e. We find that, as predicted in ref. 11, modes in the lower and upper bands are predominantly localized on the bulk and edge resonators. The modes in the centre of the bandgap, which are associated with corner charges in the case of an electrical insulator, are highly localized on the corner resonators only. Despite the finite size of the array, the ratio of the coupling rates (λ/γ) is large enough that these corner modes decay rapidly and do not overlap with each other. In Fig. 3f, we show spectra measured within the bulk bandgap of the individual corner resonators, which reveal that each corner supports only a single mid-gap mode.

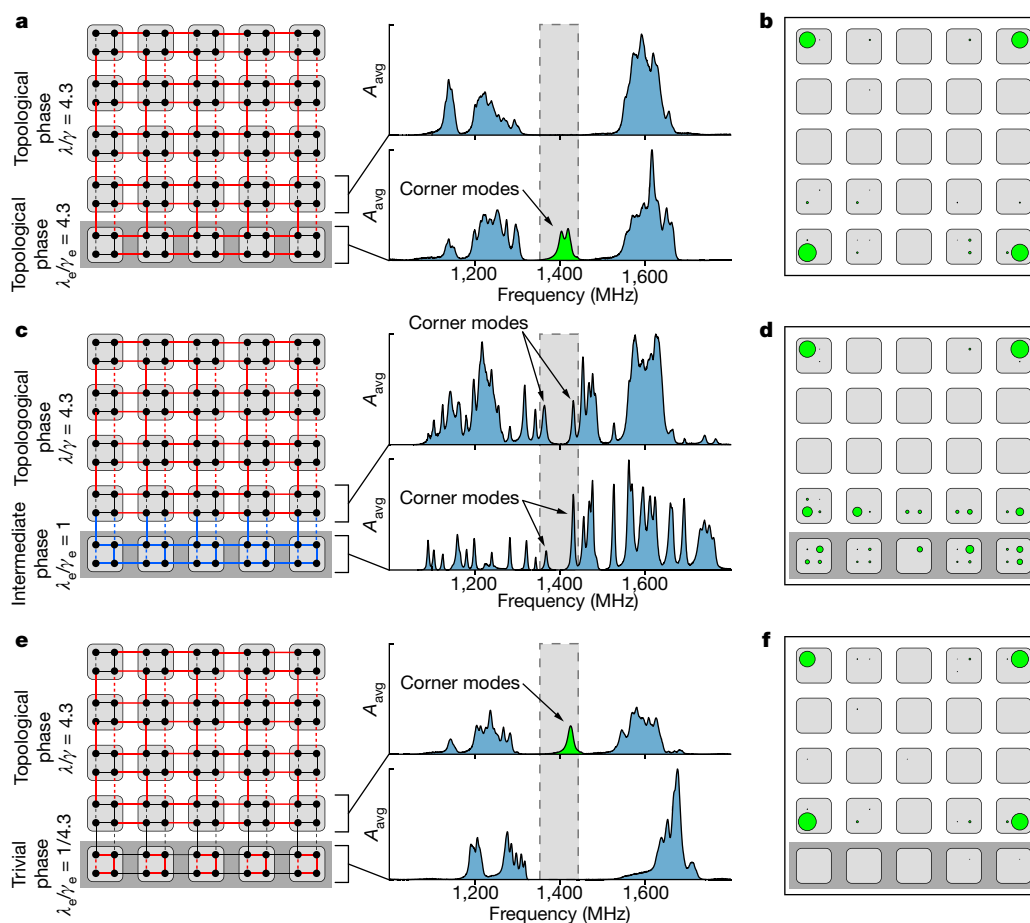


Figure 4 | Experimental test results for topologically protected corner states during edge deformation. **a**, The entire array is initially in the topological phase with $\lambda/\gamma = 4.3$. The bottom two rows of unit cells display a bandgap, with mid-gap modes—the topological corner modes—appearing only in the bottom row. **b**, Measured spatial distribution of modes within the bandgap, summed over the shaded band in **a**. **c**, The unit cells at the bottom edge are at a transition point between the topological and trivial regimes, with $\lambda_e/\gamma_e = 1$ (blue lines). The bandgap along the bottom edge narrows but remains open. Owing to the finite size of the array, the corner modes couple to each other and their degeneracy is lifted.

d, Measured spatial distribution of modes within the bandgap, summed over the shaded band in **c**. The in-gap modes are delocalized between the unit cells in the bottom two rows. **e**, The unit cells in the bottom row are brought into the trivial regime, with $\lambda_e/\gamma_e = 1/4.3$, while the rest of the array remains topological. The mid-gap modes are shifted one row up, towards the new quadrupole topological phase boundary. **f**, Measured spatial distribution of the modes corresponding to the configuration shown in **e**. The mid-gap modes are confined to the new corners of the quadrupole topological phase.

As a consequence of the disorder in the array, which breaks the chiral and reflection symmetries, the measured spectrum is asymmetric with respect to its mid-gap point. Two main sources of disorder exist (see Methods for details): (i) systematic asymmetry in the coupling rates between resonators within the array, which arises from the physical implementation of negative coupling, and (ii) random manufacturing variations in the capacitance of the discrete coupling capacitors. The main spectroscopic effect of the asymmetric coupling rates is a splitting of the lower band, which manifests in isolated plaquettes as a lifting of the degeneracy of the lower pair of modes (Fig. 2). Despite the disorder and asymmetries, we find that the robust spectral features of the quadrupole topological insulator remain; for example, the spectral bands are gapped, with only four resonances close to the mid-gap position. Furthermore, we have verified that these four mid-gap modes are tightly confined to the corners (Fig. 3e).

To demonstrate that the corner-localized modes are not the result of local effects particular to the physical edges of the array, we tune the unit cells in the lowest row of the array from the topological regime ($\gamma < \lambda$) to the trivial regime ($\gamma > \lambda$). In this experiment, the entire array is initially in the original topological phase ($\lambda/\gamma \approx 4.3$), as shown in Fig. 3. For this configuration, we plot the average absorbance spectra of the bottom two rows of unit cells separately (Fig. 4a). The spectra reveal that both rows are gapped, but the bottom row supports the

mid-gap modes, which are localized at the corners of the array (Fig. 4b). Next, we adjust the edge coupling rates in the bottom row of unit cells to be equal, that is, $\lambda_e/\gamma_e = 1$. This is achieved by changing the coupling capacitors within the network. This modification narrows the bandgap of the bottom two rows (Fig. 4c), and the two lower corner modes delocalize from the original corners into the surrounding unit cells (Fig. 4d). Owing to the finite size of the experimental array, the corner modes couple to each other at this point and their degeneracy is lifted. Finally, we make the bottom-edge unit cells trivial by setting $\lambda_e/\gamma_e \approx 1/4.3$, broadening the bandgap to its original width (Fig. 4e). Although the bottom edge of the array is now in the trivial regime, the corner modes are not destroyed, but recede to the new topological phase boundary. This experimental observation confirms that the corner modes are not a surface artefact, but are a manifestation of the bulk quadrupole topological phase. By contrast, if the corner modes were generated from local defects at the corners, or even if they arose as the end-states of edge-localized, one-dimensional topological dipole insulators, then the mid-gap modes would disappear during the edge deformation.

This work provides experimental evidence of a new family of topological phases of matter. Our metamaterial implementation of a quadrupole topological insulator confirms the existence of the theoretically predicted corner modes¹¹ and firmly establishes their origin from the

bulk quadrupole topology. This reconfigurable microwave platform can also readily support spatiotemporal modulation of both the resonance frequency and the coupling rates, enabling future experiments on dynamic topological phenomena, including pumping processes, quenches and chiral hinge modes²¹. In addition to the implementation described here, topological insulators with multipole moments could also be realized in photonic crystals, optical lattices of cold atoms¹¹ or crystalline materials³⁰, and parallel efforts have recently realized quadrupole topological insulators in electric and mechanical metamaterials^{31,32}. The stage is set for rapid advances in topological physics at both the fundamental and device levels.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 October 2017; accepted 19 January 2018.

- Zak, J. Berry's phase for energy bands in solids. *Phys. Rev. Lett.* **62**, 2747–2750 (1989).
- Vanderbilt, D. & King-Smith, R. Electric polarization as a bulk quantity and its relation to surface charge. *Phys. Rev. B* **48**, 4442–4455 (1993).
- Thouless, D., Kohmoto, M., Nightingale, M. & Den Nijs, M. Quantized Hall conductance in a two-dimensional periodic potential. *Phys. Rev. Lett.* **49**, 405–408 (1982).
- Chang, C.-Z. *et al.* Experimental observation of the quantum anomalous Hall effect in a magnetic topological insulator. *Science* **340**, 167–170 (2013).
- Kane, C. L. & Mele, E. J. Z_2 topological order and the quantum spin Hall effect. *Phys. Rev. Lett.* **95**, 146802 (2005).
- Bernevig, B. A., Hughes, T. L. & Zhang, S.-C. Quantum spin Hall effect and topological phase transition in HgTe quantum wells. *Science* **314**, 1757–1761 (2006).
- König, M. *et al.* Quantum spin Hall insulator state in HgTe quantum wells. *Science* **318**, 766–770 (2007).
- King-Smith, R. D. & Vanderbilt, D. Theory of polarization of crystalline solids. *Phys. Rev. B* **47**, 1651–1654 (1993).
- Thouless, D. J. Quantization of particle transport. *Phys. Rev. B* **27**, 6083–6087 (1983).
- Fu, L. & Kane, C. L. Time reversal polarization and a Z_2 adiabatic spin pump. *Phys. Rev. B* **74**, 195312 (2006).
- Benalcazar, W. A., Bernevig, B. A. & Hughes, T. L. Quantized electric multipole insulators. *Science* **357**, 61–66 (2017).
- Su, W. P., Schrieffer, J. R. & Heeger, A. J. Solitons in polyacetylene. *Phys. Rev. Lett.* **42**, 1698–1701 (1979).
- Hughes, T. L., Prodan, E. & Bernevig, B. A. Inversion-symmetric topological insulators. *Phys. Rev. B* **83**, 245132 (2011).
- Turner, A. M., Pollmann, F. & Berg, E. Topological phases of one-dimensional fermions: an entanglement point of view. *Phys. Rev. B* **83**, 075102 (2011).
- Leder, M. *et al.* Real-space imaging of a topologically protected edge state with ultracold atoms in an amplitude-chirped optical lattice. *Nat. Commun.* **7**, 13112 (2016).
- Meier, E. J., An, F. A. & Gadway, B. Observation of the topological soliton state in the Su-Schrieffer-Heeger model. *Nat. Commun.* **7**, 13986 (2016).
- Kraus, Y. E., Lahini, Y., Ringel, Z., Verbin, M. & Zilberberg, O. Topological states and adiabatic pumping in quasicrystals. *Phys. Rev. Lett.* **109**, 106402 (2012).
- Slobozhanyuk, A. P., Poddubny, A. N., Miroshnichenko, A. E., Belov, P. A. & Kivshar, Y. S. Subwavelength topological edge states in optically resonant dielectric structures. *Phys. Rev. Lett.* **114**, 123901 (2015).
- Blanco-Redondo, A. *et al.* Topological optical waveguiding in silicon and the transition between topological and trivial defect states. *Phys. Rev. Lett.* **116**, 163901 (2016).
- Chaunsali, R., Kim, E., Thakkar, A., Kevrekidis, P. G. & Yang, J. Demonstrating an *in situ* topological band transition in cylindrical granular chains. *Phys. Rev. Lett.* **119**, 024301 (2017).
- Benalcazar, W. A., Bernevig, B. A. & Hughes, T. L. Electric multipole moments, topological multipole moment pumping, and chiral hinge states in crystalline insulators. *Phys. Rev. B* **96**, 245115 (2017).
- Wang, Z., Chong, Y., Joannopoulos, J. D. & Soljacic, M. Observation of unidirectional backscattering-immune topological electromagnetic states. *Nature* **461**, 772–775 (2009).
- Hafezi, M., Mittal, S., Fan, J., Migdall, A. & Taylor, J. M. Imaging topological edge states in silicon photonics. *Nat. Photon.* **7**, 1001–1005 (2013).
- Rechtsman, M. C. *et al.* Photonic Floquet topological insulators. *Nature* **496**, 196–200 (2013).
- Nash, L. M. *et al.* Topological mechanics of gyroscopic metamaterials. *Proc. Natl Acad. Sci. USA* **112**, 14495–14500 (2015).
- Susstrunk, R. & Huber, S. D. Observation of phononic helical edge states in a mechanical topological insulator. *Science* **349**, 47–50 (2015).
- Ningyuan, J., Owens, C., Sommer, A., Schuster, D. & Simon, J. Time- and site-resolved dynamics in a topological circuit. *Phys. Rev. X* **5**, 021031 (2015).
- Teo, J. C. & Hughes, T. L. Existence of Majorana-fermion bound states on disclinations and the classification of topological crystalline superconductors in two dimensions. *Phys. Rev. Lett.* **111**, 047006 (2013).
- Benalcazar, W. A., Teo, J. C. & Hughes, T. L. Classification of two-dimensional topological crystalline superconductors and Majorana bound states at disclinations. *Phys. Rev. B* **89**, 224503 (2014).
- Schindler, F. *et al.* Higher-order topological insulators. Preprint at <https://arxiv.org/abs/1708.03636> (2017).
- Imhof, S. *et al.* Topoelectrical circuit realization of topological corner modes. Preprint at <https://arxiv.org/abs/1708.03647> (2017).
- Serra-Garcia, M. *et al.* Observation of a phononic quadrupole topological insulator. *Nature* **555**, <https://doi.org/10.1038/nature25156> (2018).

Acknowledgements We would like to thank J. T. Bernhard for access to the resources at the UIUC Electromagnetics Laboratory. This project was supported by the US National Science Foundation (NSF) through the Emerging Frontiers in Research and Innovation (EFRI) grant EFMA-1627184. C.W.P. acknowledges support from an NSF Graduate Research Fellowship. G.B. acknowledges support from the US Office of Naval Research (ONR) through a Director for Research Early Career Grant. W.A.B. and T.L.H. thank the US NSF for grant DMR-1351895.

Author Contributions C.W.P. designed the microwave quadrupole topological insulator, performed the microwave simulations and experimental measurements and produced the experimental figures. W.A.B. guided the topological insulator design and performed the theoretical calculations. T.L.H. and G.B. supervised all aspects of the project. All authors jointly wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to G.B. (bahl@illinois.edu).

Reviewer Information *Nature* thanks Y. Kivshar and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Spectrum and eigenmode measurements. We measure the power absorbance spectrum at each resonator in the tested networks by means of one-port reflection (S_{11}) measurements using a microwave network analyser (Keysight E5063A). The reflection probe is composed of a 50 Ω coaxial cable terminated in a 0.1 pF capacitor, which is connected to each resonator at an anti-node. Owing to the low probe capacitance, the measured linewidths are dominated by the intrinsic losses of each resonator.

The absorbance of each resonator is calculated as $A = 1 - |S_{11}|^2$. We also define the average absorbance for an array of N resonators as $A_{\text{avg}} = \frac{1}{N} \sum_n A_n$, where A_n is the absorbance of the n th resonator. The average background absorbance contribution from the network analyser probe is evaluated far from any modes and removed during the measurements of A_n .

The eigenmodes of the unit cell and 2×2 array (Fig. 2) are also measured with a microwave network analyser by means of two-port transmission (S_{21}) measurements. The measurements are performed using a pair of probes with the specifications described above, with one probe providing the stimulus and the other measuring the response. Thus, the S_{21} transfer function at the resonance frequency represents a direct measurement of the amplitude and phase response of the corresponding eigenmode.

Design of the quadrupole topological insulator lattice. Each unit cell is fabricated individually on a Rogers RT/duroid 5880 substrate, with 35- μm -thick copper coating on each side. An approximate transmission line representation of our resonator is shown in Extended Data Fig. 1a. The resonator is H-shaped, with sections of approximately the same length (1.5 cm), width (0.1 cm) and characteristic impedance (about 110 Ω). This resonator design leads to an unloaded resonance frequency of 2.1 GHz in the transmission line model (the measured resonance frequency of a fabricated resonator is 2.08 GHz). Although there are losses in both the dielectric substrate and the copper conductor, as well as negligible radiative losses, these are small and do not affect the underlying topology.

To create a unit cell, four microstrip resonators are capacitively coupled as shown in Extended Data Fig. 1b. Each coupling is created using two 0.2 pF capacitors in series, resulting in a total coupling capacitance of 0.1 pF (related to the coupling parameter γ) between resonators within the unit cell. Negative coupling is realized by connecting R1 to the opposite-phase anti-node, R4.

The connections between unit cells are detailed in Extended Data Fig. 1c. Each capacitive coupling is formed using two 2 pF capacitors in series, resulting in a total inter-cell coupling capacitance (related to the coupling parameter λ) of 1 pF. The average coupling rates γ and λ are extracted from the measured data in the limits $\lambda \rightarrow 0$ and $\gamma \rightarrow 0$, respectively (Fig. 2). We find that the ratio of the frequency separations between the degenerate mode pairs in these isolated intra-unit-cell and inter-unit-cell cases is on average approximately 4.3, which implies a coupling rate ratio of $\lambda/\gamma \approx 4.3$.

Comparison of unit cells threaded with different fluxes. A unit cell of our quadrupole topological insulator is a square composed of four resonators threaded with π flux, as illustrated in Fig. 2a. In Fig. 2b, we show the measured eigenmodes of a single unit cell, which match well with the theoretically predicted modes. Here we discuss all four eigenmodes of this system and establish its differences from unit cells threaded with 0 flux. We find that without flux the modes differ substantially in their spatial distributions but their energy spectra can be similar if C_4 symmetry is broken.

The calculated energy spectrum and eigenmodes of a unit cell with π flux and equal coupling rates γ are shown in Extended Data Fig. 2a, along with a graphical representation of the Hamiltonian

$$H = \begin{pmatrix} 0 & 0 & \gamma & -\gamma \\ 0 & 0 & \gamma & \gamma \\ \gamma & \gamma & 0 & 0 \\ -\gamma & \gamma & 0 & 0 \end{pmatrix} \quad (1)$$

As noted in the main manuscript, there are two pairs of degenerate eigenmodes. These modes can be described by the orthonormal basis vectors

$$\begin{aligned} \mathbf{u}_1 &= (1/2 \quad -1/2 \quad 0 \quad \sqrt{2}/2) \\ \mathbf{u}_2 &= (-1/2 \quad -1/2 \quad \sqrt{2}/2 \quad 0) \\ \mathbf{u}_3 &= (-1/2 \quad 1/2 \quad 0 \quad \sqrt{2}/2) \\ \mathbf{u}_4 &= (1/2 \quad 1/2 \quad \sqrt{2}/2 \quad 0) \end{aligned} \quad (2)$$

where the elements of each vector are expressed relative to the complex amplitudes of the resonators R1, R2, R3 and R4. \mathbf{u}_1 and \mathbf{u}_2 represent the degenerate pair of

lower-energy modes, and \mathbf{u}_3 and \mathbf{u}_4 describe the degenerate pair of higher-energy modes. In Fig. 2b we show the measurement results for modes \mathbf{u}_1 and \mathbf{u}_2 . Owing to the destructive interference arising from the π flux within the plaquette, when one resonator is excited (here R3 or R4), the resonator in the opposite corner is not excited. This property leads to the characteristic modes of the unit cell. We also find that the location of the negative coupling affects the relative phase between the resonators, leading to the opposite relative phase between resonators with and without negative coupling.

We compare the above case with that of an identical unit cell with 0 flux, shown in Extended Data Fig. 2b and described by the Hamiltonian

$$H = \begin{pmatrix} 0 & 0 & \gamma & \gamma \\ 0 & 0 & \gamma & \gamma \\ \gamma & \gamma & 0 & 0 \\ \gamma & \gamma & 0 & 0 \end{pmatrix} \quad (3)$$

These modes can be described by the orthonormal basis vectors

$$\begin{aligned} \mathbf{v}_1 &= (1/2 \quad 1/2 \quad -1/2 \quad -1/2) \\ \mathbf{v}_2 &= (0 \quad 0 \quad -\sqrt{2}/2 \quad \sqrt{2}/2) \\ \mathbf{v}_3 &= (\sqrt{2}/2 \quad -\sqrt{2}/2 \quad 0 \quad 0) \\ \mathbf{v}_4 &= (1/2 \quad 1/2 \quad 1/2 \quad 1/2) \end{aligned} \quad (4)$$

In this unit cell, only modes \mathbf{v}_2 and \mathbf{v}_3 are degenerate, whereas \mathbf{v}_1 has lower energy and \mathbf{v}_4 has higher energy than \mathbf{v}_2 and \mathbf{v}_3 . Because there is 0 flux threading the unit cell, when one resonator is excited, the resonator in the opposite corner is always excited as well.

Although a unit cell with 0 flux and identical horizontal and vertical coupling rates ($\gamma_x = \gamma_y$) is not gapped, a bandgap can be opened by setting $\gamma_x > \gamma_y$ (that is, by breaking C_4 symmetry). The energy spectrum and eigenmodes calculated for this case are shown in Extended Data Fig. 2c, along with a graphical representation of the Hamiltonian

$$H = \begin{pmatrix} 0 & 0 & \gamma_x & \gamma_y \\ 0 & 0 & \gamma_y & \gamma_x \\ \gamma_x & \gamma_y & 0 & 0 \\ \gamma_y & \gamma_x & 0 & 0 \end{pmatrix} \quad (5)$$

The modes can be described by the orthonormal basis vectors

$$\begin{aligned} \mathbf{w}_1 &= (1/2 \quad 1/2 \quad -1/2 \quad -1/2) \\ \mathbf{w}_2 &= (1/2 \quad -1/2 \quad 1/2 \quad -1/2) \\ \mathbf{w}_3 &= (1/2 \quad -1/2 \quad -1/2 \quad 1/2) \\ \mathbf{w}_4 &= (1/2 \quad 1/2 \quad 1/2 \quad 1/2) \end{aligned} \quad (6)$$

Although none of these modes is degenerate, the lower pair (and upper pair) can be brought arbitrarily close for a large ratio γ_x/γ_y . However, the spatial distribution of these eigenmodes clearly differs from that of a unit cell with π flux, since all four resonators are excited equally in each mode.

Systematic and random disorder caused by capacitive loading. The coupling capacitors also capacitively load the resonators, increasing their effective length and therefore reducing the resonance frequency. For bulk resonators, the capacitive loading is similar and does not affect the bulk spectral characteristics. However, the reduced capacitive loading for edge and corner resonators is compensated (to match the bulk loading) by adding a capacitance-to-ground of 0.6 pF and 1.2 pF to the edge and corner resonators, respectively. Small differences in the capacitive loading of the resonators inside our quadrupole topological insulator array imply disorder in both the resonance frequencies and the coupling rates.

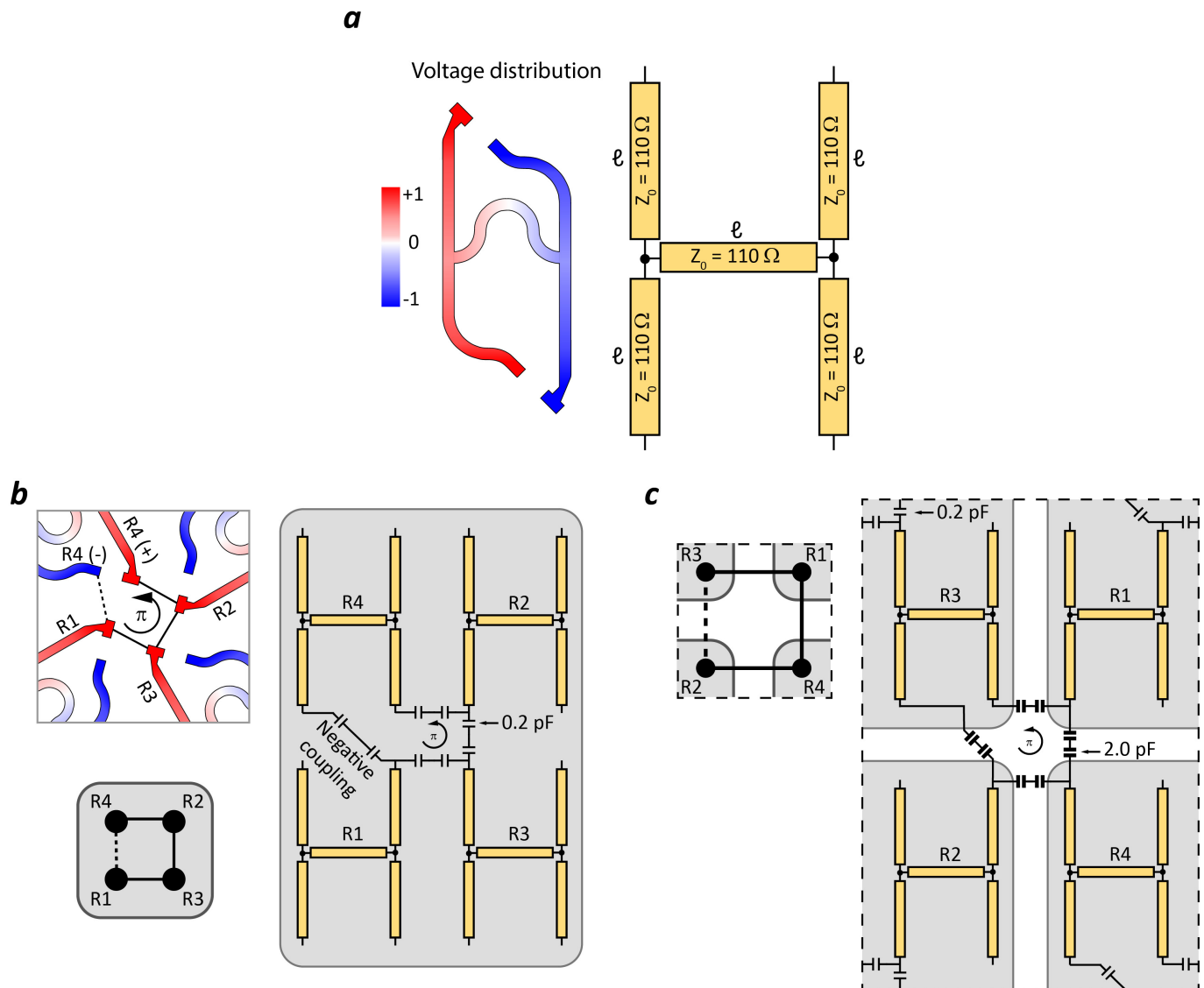
The impact of this disorder is seen in the measured eigenmodes in the limits $\gamma \rightarrow 0$ and $\lambda \rightarrow 0$ (Fig. 2). In the full array, such disorder also results in splitting of the lower bulk band (Fig. 3b). To understand how these differences in capacitive loading arise, we examine two representative cases of resonators (Extended Data Fig. 3) loaded with identical total capacitance, but distinct spatial distributions.

In Extended Data Fig. 3a, we show the case where both capacitors are on the same arm of the resonator, as is the case for the intra-unit-cell coupling of resonators R1, R2 and R3 (Extended Data Fig. 1b) and the inter-unit-cell coupling of resonators R1, R2 and R4 (Extended Data Fig. 1c). The addition of a 2 pF capacitance-to-ground on one arm of the resonator causes a frequency shift to

1.4 GHz (compared with 2.1 GHz for the unloaded resonator). In Extended Data Fig. 3b, we examine the case where the same 2 pF capacitance is distributed to the two arms of the resonator, as is the case for the intra-unit-cell coupling of resonator R4 and for the inter-unit-cell coupling of resonator R3. Here, the resonance frequency shifts to 1.3 GHz; that is, it is lower than when both resonators are on the same arm. Although the total capacitance on each resonator is the same, these representative cases illustrate that the spatial distribution of capacitors affects the degree of capacitive loading on the resonator. Thus, systematic disorder in the capacitive loading of each resonator, which affects the coupling rate between resonators and their resonance frequencies, arises throughout our array.

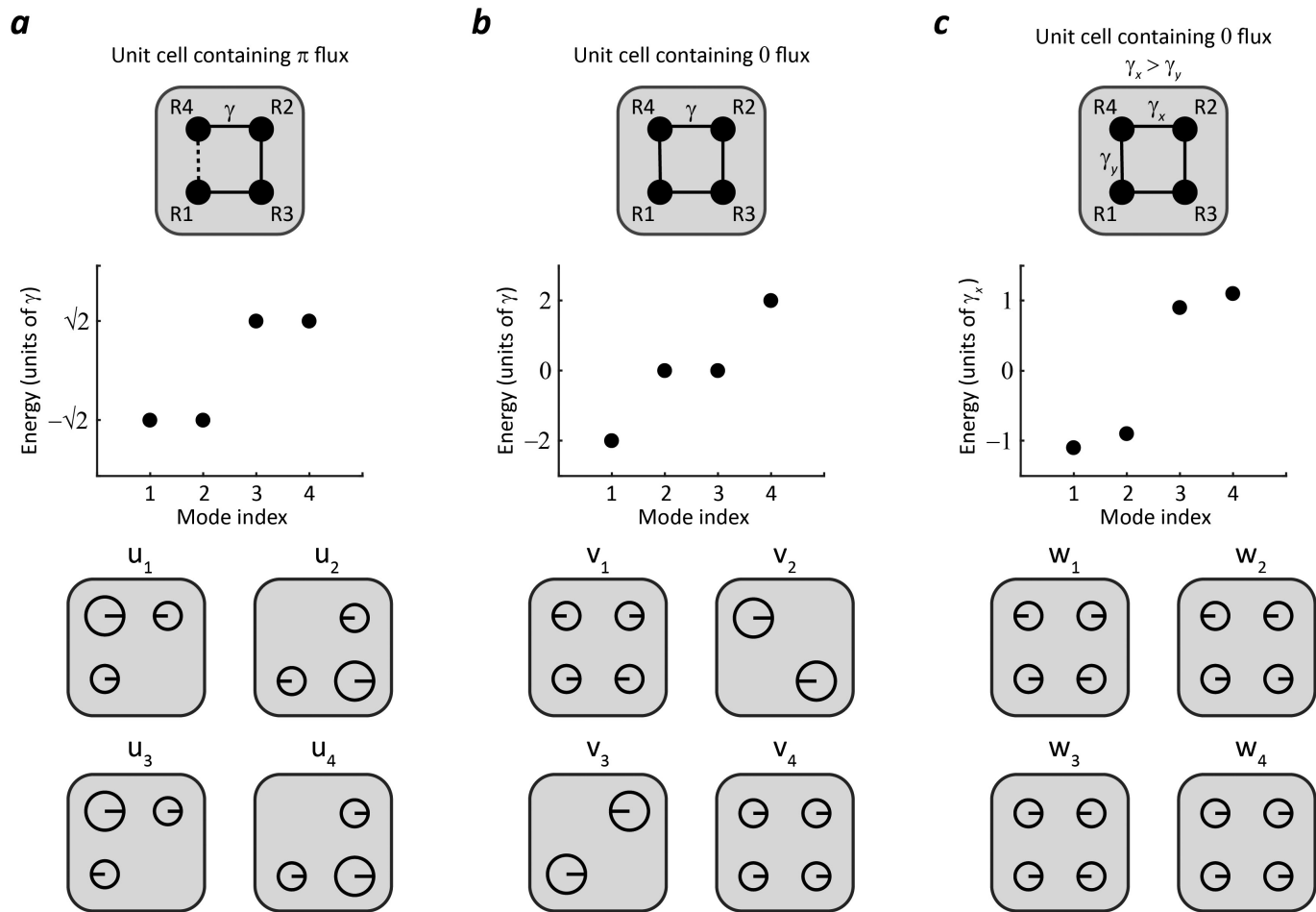
In addition, the system exhibits random disorder due to variations in the manufacturing of its discrete components. Specifically, the 0.2 pF capacitors have a tolerance of ± 0.05 pF and the 2 pF capacitors have a tolerance of ± 0.1 pF. Despite this difference, the ratio λ/γ remains much larger than 1 throughout the array, so the bandgap remains open and the system remains firmly in the topological phase. A final source of disorder in the spectrum is the increase in the capacitive coupling rate with increasing frequency. Because of this effect, the lower bulk band is broader in frequency than the upper band.

Data availability. The data that support the findings of this study are available from the corresponding author on reasonable request.



Extended Data Figure 1 | Transmission line model. **a**, Transmission line model of an individual microstrip resonator. Each section has approximately the same length $\ell = 1.5\text{ cm}$ and the same characteristic impedance $Z_0 = 110\ \Omega$, which give a fundamental resonance frequency of 2.1 GHz. **b**, Resonators coupled within the unit cell by two 0.2 pF

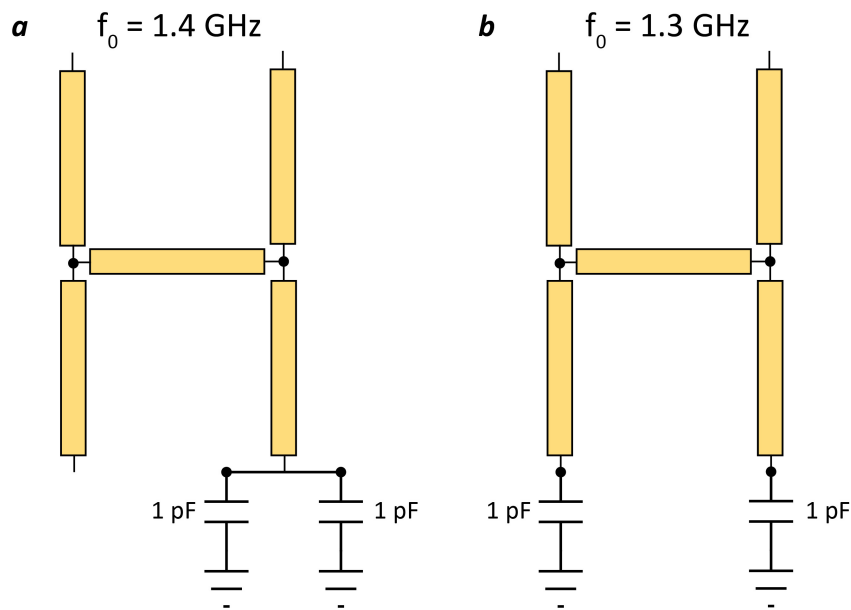
capacitors in series. The capacitors linking R1 and R4 are connected to the out-of-phase anti-node of R4, creating π -flux threading in the plaquette. **c**, Coupling of resonators in different unit cells by two 2-pF capacitors in series. The capacitors between R2 and R3 are connected to the out-of-phase anti-node of R3 to produce the required π flux.



Extended Data Figure 2 | Comparison of unit cells threaded with π and 0 flux. **a**, Energy spectrum and eigenmodes of a unit cell with π flux.

b, Energy spectrum and eigenmodes of a unit cell with 0 flux and $\gamma_x = \gamma_y$.

c, Energy spectrum and eigenmodes of a unit cell with 0 flux and unequal coupling rates $\gamma_x > \gamma_y$. The energy separation between the lower two (and upper two) modes is proportional to γ_y .



Extended Data Figure 3 | Comparison of resonators loaded with equal total capacitance. The resonance frequencies are calculated from simulations using Keysight ADS. **a**, A resonator with 2-pF loading on a single arm. The resonance frequency is shifted from 2.1 GHz to 1.4 GHz because of the loading. This situation corresponds to that of the intra-unit-cell coupling of resonators R1, R2 and R3 (Extended Data Fig. 1b) and

the inter-unit-cell coupling of resonators R1, R2 and R4 (Extended Data Fig. 1c). **b**, A resonator with 2-pF loading distributed to two opposite-polarity arms. The resonance frequency is shifted from 2.1 GHz to 1.3 GHz owing to the loading. This situation is representative of the intra-unit-cell coupling of resonator R4 and the inter-unit-cell coupling of resonator R3.

High-resolution magnetic resonance spectroscopy using a solid-state spin sensor

David R. Glenn^{1*}, Dominik B. Bucher^{1,2*}, Junghyun Lee³, Mikhail D. Lukin¹, Hongkun Park^{1,4} & Ronald L. Walsworth^{1,2}

Quantum systems that consist of solid-state electronic spins can be sensitive detectors of nuclear magnetic resonance (NMR) signals, particularly from very small samples. For example, nitrogen-vacancy centres in diamond have been used to record NMR signals from nanometre-scale samples^{1–3}, with sensitivity sufficient to detect the magnetic field produced by a single protein⁴. However, the best reported spectral resolution for NMR of molecules using nitrogen-vacancy centres is about 100 hertz⁵. This is insufficient to resolve the key spectral identifiers of molecular structure that are critical to NMR applications in chemistry, structural biology and materials research, such as scalar couplings (which require a resolution of less than ten hertz⁶) and small chemical shifts (which require a resolution of around one part per million of the nuclear Larmor frequency). Conventional, inductively detected NMR can provide the necessary high spectral resolution, but its limited sensitivity typically requires millimetre-scale samples, precluding applications that involve smaller samples, such as picolitre-volume chemical analysis or correlated optical and NMR microscopy. Here we demonstrate a measurement technique that uses a solid-state spin sensor (a magnetometer) consisting of an ensemble of nitrogen-vacancy centres in combination with a narrowband synchronized readout protocol^{7–9} to obtain NMR spectral resolution of about one hertz. We use this technique to observe NMR scalar couplings in a micrometre-scale sample volume of approximately ten picolitres. We also use the ensemble of nitrogen-vacancy centres to apply NMR to thermally polarized nuclear spins and resolve chemical-shift spectra from small molecules. Our technique enables analytical NMR spectroscopy at the scale of single cells.

In recent years, optically probed nitrogen-vacancy (NV) centres in diamond have become the leading modality for magnetic sensing at short length scales (nanometres to micrometres) under ambient conditions, with wide-ranging application in the physical and life sciences¹⁰. So far, two key challenges have limited the spectral resolution of NMR detection using NV centres (NV-NMR). First, NV-NMR has been demonstrated only in nanometre-scale measurement volumes, in which the thermal (Boltzmann) spin polarization is too small to observe, but statistical fluctuations in the spin polarization¹¹ are large and easily detected¹². However, such nanometre-scale samples have short, diffusion-limited spin-noise correlation times that give rise to broad NMR spectral lines (typically much broader than 1 kHz)¹³. Second, the interrogation duration for NV-NMR detection techniques has generally been limited by the spin-state lifetime of the NV centre ($T_1 \approx 3$ ms), which is orders of magnitude shorter than the coherence times of nuclear spins in bulk liquid samples ($T_2 \approx 1$ s). Recent studies have shown that quantum memories can greatly extend the useful NV spin lifetime^{14,15}. Nonetheless, such techniques have unfavourable sensitivity scaling with spectral resolution, $\eta \propto (\Delta f)^{-1/2}$, because the NV probe must be in a non-interacting state while the memory is active⁵, and are still fundamentally limited by spin diffusion in the sample when applied at the nanometre scale. Here, we address these challenges

and achieve an NV-NMR spectral resolution of around 1 Hz by: (i) probing micrometre-scale measurement volumes to obtain a signal that is dominated by the thermal spin polarization, which is not limited by diffusion; and (ii) using a synchronized readout protocol to sense NMR signals coherently for an arbitrary duration (up to around 10^3 s).

We first present a sensitive ($\eta_B = 30$ pT Hz^{-1/2}; Extended Data Fig. 1) NV-ensemble magnetometer, which is designed to detect NMR signals from the thermal spin polarization of a micrometre-scale sample. The sensor volume consists of the overlap region between a 13- μ m-thick NV-doped layer at the diamond surface and a 20- μ m-diameter optical excitation beam (Fig. 1a). To motivate this sensor geometry, we consider a single NV centre, located at a depth d_{NV} below the diamond surface, and a liquid sample that consists of a half-space of Larmor-precessing spins above the surface. For the NV-NMR signal due to the thermal spin polarization of the sample to dominate the noise from magnetic fluctuations of the sample, d_{NV} must be greater than about 3 μ m (Supplementary Note 2). On the other hand, the effective measurement volume for the thermally polarized sample increases as d_{NV}^3 (Fig. 1a inset). Our sensor design, with a mean NV-centre depth of $\bar{d}_{\text{NV}} = 6.5$ μ m, realizes a compromise between (i) being insensitive to broadband magnetic noise caused by rapid near-surface spin fluctuations, and (ii) maintaining a small effective NMR measurement volume of about 10 picolitres (Extended Data Fig. 3).

Interrogation of the NV-ensemble sensor using a coherently averaged synchronized readout (CASR) scheme⁷ provides the spectral selectivity that is needed for molecular NMR spectroscopy. The CASR protocol (Fig. 1b) consists of concatenated NV magnetometry pulse sequences, interspersed with projective NV spin-state readouts, all synchronized to an external clock. In the limit of weak coupling between the NV centres and the signal source, the NV measurement back-action is small and does not lead to direct dephasing of the spins in the sample (Extended Data Fig. 4). The detector line width is then limited only by technical effects (such as gradients in the bias magnetic field B_0) and the stability of the clock. Crucially, because our sensor is optimized to detect the thermal spin polarization, the phase of the NMR signal can be made identical over repetitions of the CASR protocol by the application of an initial $\pi/2$ pulse to the nuclear spins at time $t = 0$, which enables coherent signal averaging. (Incoherent averaging of synchronized readout measurements is also possible, but leads to poor sensitivity scaling; see Supplementary Note 5.) To characterize the spectral-resolution limit of the synchronized readout pulse sequences due to our timing source, we applied an oscillating magnetic signal consisting of three closely spaced frequencies using a nearby coil antenna, and measured it using CASR. We observed line widths of 0.4 mHz (Extended Data Fig. 5), which is several orders of magnitude better resolution than necessary for identifying molecular NMR signatures such as scalar couplings (J -couplings) and chemical shifts.

We performed initial CASR NMR measurements using a sample of glycerol ($\text{C}_3\text{H}_8\text{O}_3$) molecules (Fig. 2a). The NV-ensemble sensor was placed in a cuvette filled with glycerol and aligned in the bias field

¹Department of Physics, Harvard University, Cambridge, Massachusetts, USA. ²Harvard-Smithsonian Centre for Astrophysics, Cambridge, Massachusetts, USA. ³Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. ⁴Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts, USA.

*These authors contributed equally to this work.

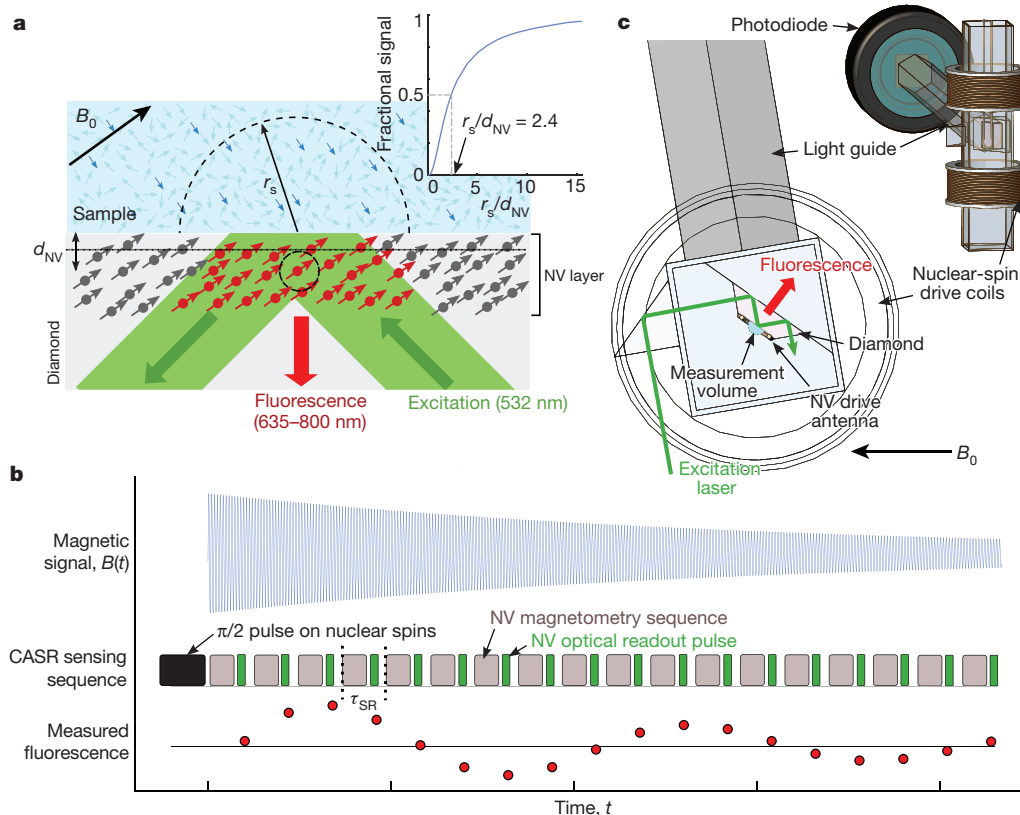


Figure 1 | NV-ensemble sensor for CASR NMR. **a**, Geometry of the NV-ensemble sensor. The sensor consists of an NV-centre layer approximately $13 \mu\text{m}$ thick at the surface of a diamond chip, which is probed by a green laser (532-nm excitation wavelength) with a beam diameter of about $20 \mu\text{m}$. The sensor detects NMR from the thermally polarized fraction (dark blue arrows) of the total nuclear spin population of the sample (light blue arrows). Sample spins are resonantly driven and precess around the static bias field B_0 . NV centres at depths of $d_{NV} > 3 \mu\text{m}$ (dotted horizontal line) are primarily sensitive to the thermal spin signal; shallower NV centres have signals that are dominated by statistical spin fluctuations. For an NV centre at depth d_{NV} (such as that shown in the dashed circle), half of the NMR signal is due to sample spins in a hemisphere of radius $r_s < 2.4d_{NV}$ (for r_s indicated by the grey semicircle), which defines the effective sensing volume for that NV centre. The inset shows a numerical calculation of the fractional NMR signal amplitude that originates in a hemisphere of radius r_s , normalized to the full NMR signal amplitude obtained when r_s becomes much larger than d_{NV} . **b**, Numerical simulation of CASR detection of an NMR FNP signal. The

($B_0 = 88 \text{ mT}$) of a feedback-stabilized electromagnet. A resonant $\pi/2$ pulse was applied to tip the thermally polarized proton spins in the sample into the transverse plane of the Bloch sphere. The free-nuclear-precession (FNP) signal of the proton (equivalent to the free-induction-decay signal in conventional NMR) was then measured using a CASR sequence. Near the end of this sequence, after the spins in the sample were fully dephased, we used a coil antenna to apply a calibrated oscillating-magnetic-field pulse. A comparison of the integrated peak intensities of the glycerol NMR and the coil pulse signals in the CASR amplitude spectrum (Fig. 2a inset) yielded an initial glycerol proton FNP amplitude of $95 \pm 8 \text{ pT}$ (1σ , for $n = 3$ measurements), which is approximately consistent with calculations (Supplementary Note 2). To exclude the possibility of a spurious detection associated with room noise or sensor imperfections, we swept B_0 over a range of 0.02 mT (in steps of approximately 0.005 mT) and repeated the CASR FNP experiment at each value. A linear fit to the resulting NMR frequencies gives the correct value for the proton gyromagnetic ratio (Fig. 2b).

To assess the spectral resolution limits of NMR detection using CASR spectroscopy, we measured a sample of pure water ($T_2, T_1 > 2 \text{ s}$; ref. 16). The resulting NMR signal line width (quoted as the full-width

at half-maximum, FWHM) is $9 \pm 1 \text{ Hz}$ (Fig. 2c), which we attribute to micrometre-scale magnetic gradients from susceptibility differences between sensor components (Extended Data Fig. 6). Gradient-induced spectral broadening is commonly observed in submicrolitre-volume NMR spectroscopy with microcoils^{17,18} and can be mitigated by improving susceptibility matching in the sensor design¹⁹. Applying π pulses to the protons to refocus the gradient-induced dephasing (that is, CASR spin-echo) narrowed the NMR signal line width to 2.8 Hz (FWHM; Fig. 2c), in agreement with the distribution of temporal fluctuations in B_0 that were recorded during the experiment (Extended Data Fig. 2). We experimentally investigated sensor-induced sample dephasing due to spatially inhomogeneous interactions with the NV spins and/or the microwaves used to manipulate the NV centres; neither effect contributed substantially to the measured NMR signal line width (Supplementary Data 7).

To illustrate the applicability of CASR to molecular NMR spectroscopy, we acquired liquid-state FNP signals from picolitre-volume samples of three molecules: (i) trimethyl phosphate (TMP) $[\text{PO}(\text{OCH}_3)_3]$, which is known to have a J -coupling of $J(\text{P}, \text{H}) \approx 11 \text{ Hz}$ between the methyl protons and the central ^{31}P

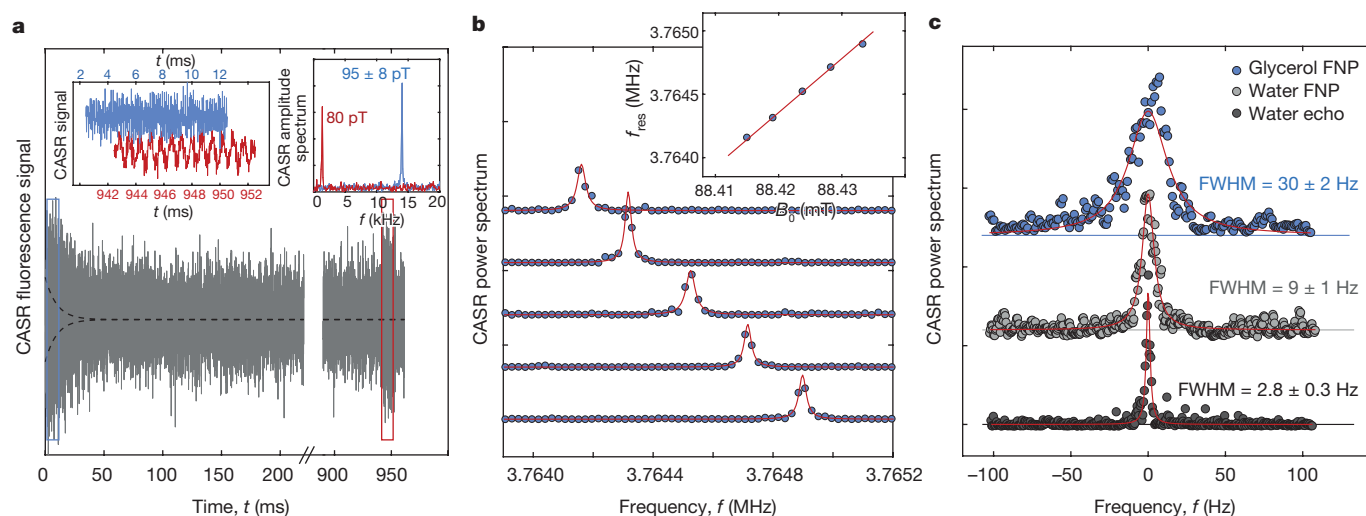


Figure 2 | NMR detection using CASR. **a**, CASR time-series signal (grey trace) produced by NMR FNP of glycerol proton spins above the diamond, with decay time $T_2^* = 10 \pm 1$ ms (the black dashed line shows a noise-subtracted fit of the exponential decay envelope). A calibrated (80-pT) amplitude; red peak in the right inset) magnetic field from a coil antenna is turned on at $t = 940$ ms. A comparison of FNP (blue box, blue trace in the left inset) and antenna (red box, red trace in the left inset) signals in the frequency domain (right inset) yields an initial FNP amplitude (blue peak in the right inset) of 95 ± 8 pT. Total signal averaging time was 7.2×10^4 s. **b**, Power spectra of proton NMR signals obtained from glycerol CASR FNP data (blue circles) for varying B_0 , fitted to Lorentzian line shapes (solid red lines). A linear fit of the NMR resonance frequency

f_{res} versus B_0 (inset) gives the correct proton gyromagnetic ratio, $\gamma_p = 42.574 \pm 0.002$ MHz T^{-1} . The signal averaging time was 2.8×10^3 s per trace. **c**, Power spectra of CASR FNP measured from protons in glycerol (blue circles) and pure water (grey circles), as well as CASR spin-echo measured from pure water (black circles); spectra are offset for clarity. The spectral resolution obtained with CASR FNP of glycerol is 30 ± 2 Hz (FWHM), as determined by least-squares fitting to a Lorentzian line shape (red line). The spectral resolution obtained from pure water is 9 ± 1 Hz (FWHM) with CASR FNP and 2.8 ± 0.3 Hz (FWHM) with CASR spin-echo. The signal averaging times were 7.2×10^4 s (glycerol FNP), 3.1×10^4 s (water FNP) and 3.9×10^4 s (water spin-echo).

nuclear spin²⁰; (ii) xylene [(CH₃)₂C₆H₄], which has substantially different chemical shifts (about 5 p.p.m.) associated with the carbon ring structure and with the satellite methyl groups; and (iii) ethyl formate [HCOOC₂H₅], which has three different chemical shifts and two unequal J -couplings. The CASR NMR spectrum for TMP (Fig. 3a) shows clearly resolved peaks due to the J -coupling between the protons and the spin-1/2 ³¹P nucleus, with a splitting

of $\Delta f \approx 13 \pm 1$ Hz. The CASR NMR spectrum for xylene (Fig. 3b) shows peaks split by 5.3 ± 0.5 p.p.m. (equivalent to a difference in chemical-shift frequency of $\Delta f_{\text{CS}} \approx 20 \pm 2$ Hz at our bias field), consistent with a previously reported²¹ value for the chemical shift. The observed peak intensity ratio of about 2.2:1 in the xylene NMR power spectrum corresponds to the relative nuclear abundance of 6:4, with the protons in high-electron-density methyl groups shifted to lower frequency. The

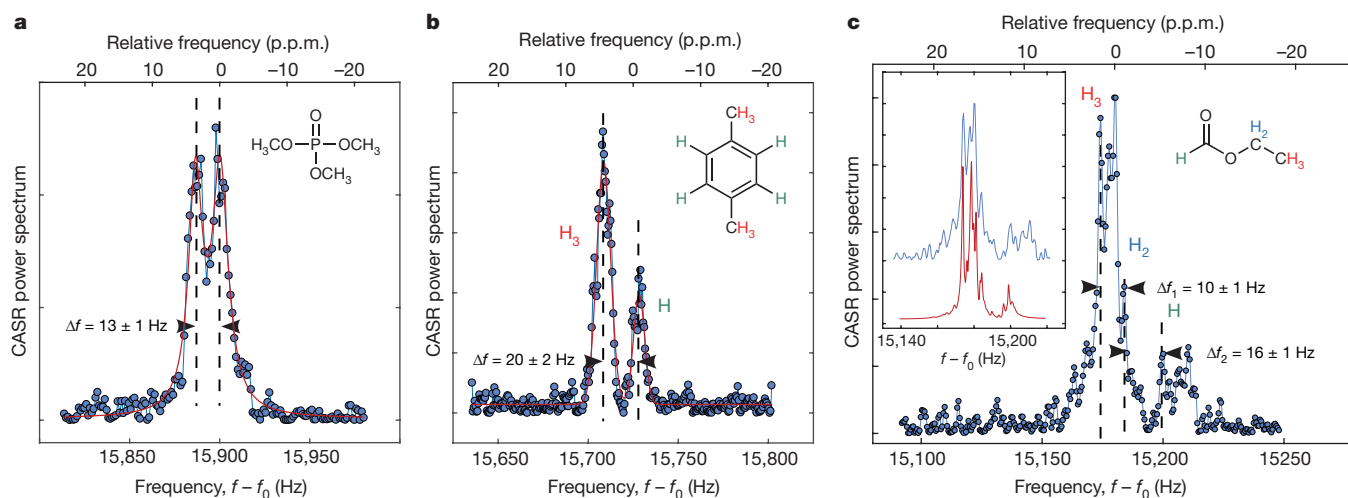


Figure 3 | CASR-detected molecular NMR spectra. **a**, CASR FNP spectrum of trimethyl phosphate (blue circles; see inset for chemical structure). The fit (solid red line) indicates a splitting of $\Delta f = 13 \pm 1$ Hz (indicated by the vertical dashed lines and arrows) due to scalar coupling (J -coupling) between the central ³¹P nucleus and the methyl protons. The top axis shows the relative frequency $(f - f_{\text{mid}})/f_0$, where $f_{\text{mid}} - f_0 = 15,900$ Hz is defined as the midpoint of the spectrum and $f_0 = 3.74$ MHz is the central frequency of the CASR sequence. The signal averaging time was 3.3×10^4 s. **b**, CASR FNP spectrum of xylene (blue circles; see inset for chemical structure). The relative peak heights obtained from the fit (solid red lines) are due to the relative abundances of

CH (left peak) and CH₃ (right peak) protons in the molecule. The splitting of 5.3 ± 0.5 p.p.m. (or $\Delta f = 20 \pm 2$ Hz) is the result of chemical shifts associated with the two proton positions (labelled H and H₃). The signal averaging time was 3.3×10^4 s. **c**, Measured CASR FNP spectrum of ethyl formate (blue circles; see right inset for chemical structure). Individual peaks are identified with frequency differences of 2.7 ± 0.3 p.p.m. (or $\Delta f_1 = 10 \pm 1$ Hz) and 2.7 ± 0.3 p.p.m. (or $\Delta f_2 = 16 \pm 1$ Hz), which correspond to the chemical shifts of the three proton groups (labelled H, H₂ and H₃). The inset shows a comparison between the measured (blue) and calculated (red) ethyl formate FNP spectra (offset for clarity). The signal averaging time was 3.5×10^4 s.

CASR NMR spectrum for ethyl formate (Fig. 3c) is comparatively complex, owing to multiple chemical-shift constants and J -couplings that are all in the range 3–25 Hz. Because of the strong couplings, second-order effects cause resonance lines to be split at higher multiplicities than typically observed in high-field NMR, bringing some spectral features close to our noise floor. Nevertheless, the CASR measurement shows good qualitative agreement with a spectrum calculated using molecular parameters estimated from standard (sample size of roughly 1 ml) high-field NMR (Extended Data Fig. 7). These measurements constitute a demonstration of NMR at picolitre-scale volumes with a spectral resolution that is sufficient to resolve molecular J -couplings and chemical shifts, and of NV-NMR for a thermally polarized sample.

CASR spectroscopy is not subject to the line-width limitations that are usually associated with the small values of T_1 and T_2 for NV spins. The NV-ensemble sensor detects thermal spin polarization without broadening, because the sample correlation time is not limited by diffusion. The combined technique thus provides a spectral resolution that is about two orders of magnitude better than that demonstrated previously with NV-NMR of molecules in liquid samples. Increasing B_0 from 88 mT to 3 T should improve the proton-number sensitivity and the chemical-shift resolution of the CASR sensor by more than an order of magnitude, while mitigating the spectral complexity that is associated with strong J -couplings at low field. Integration with microfluidics will provide more optimal matching of sample and measurement volumes to enable efficient NMR spectroscopy of mass-limited samples. It might also be possible to realize large increases in the sensitivity of CASR spectroscopy by using techniques proposed for diamond-based dynamic nuclear polarization^{22,23}.

NV-ensemble CASR is well suited to micrometre-scale (picolitre-scale) measurement volumes. With further improvements in sensitivity, the technique could enable NMR spectroscopy of small molecules and proteins at the single-cell level²⁴—a longstanding scientific goal that has yet to be realized for eukaryotic cells smaller than *Xenopus laevis* ova²⁵ or the L7 neuron of *Aplysia californica*²⁶. Potential applications include NMR studies of single-cell metabolomics²⁷ and NMR fingerprinting of heterogeneous protein expression in tumour cells²⁸. In addition, the diamond chip is compatible with correlative optical microscopy²⁹ with a field of view of about 1 mm², and the NV-ensemble CASR sensing volume may be positioned anywhere on the chip by adjusting the alignment of the laser excitation. These capabilities suggest applications to networks of cells, such as spatially resolved NMR of signalling molecules in bacterial biofilms³⁰. Finally, although this work has focused on high-resolution NMR at the micrometre-scale, CASR is equally applicable to bulk, millimetre-sized NV-ensemble detectors (Supplementary Note 9), which could provide NMR concentration sensitivity comparable to that obtained using inductive microcoils in microlitre-scale measurement volumes (Extended Data Fig. 8). Bulk NV-ensemble sensors should be amenable to parallel operation, using an array of diamond chips with independent (cross-talk-free) optical readouts for each. This opens up the possibility of parallelized, high-throughput analytical NMR spectroscopy for concentration-limited samples.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 July 2017; accepted 3 January 2018.

1. Mamin, H. J. *et al.* Nanoscale nuclear magnetic resonance with a nitrogen-vacancy spin sensor. *Science* **339**, 557–560 (2013).
2. Staudacher, T. *et al.* Nuclear magnetic resonance spectroscopy on a (5-nanometer)³ sample volume. *Science* **339**, 561–563 (2013).
3. DeVience, S. J. *et al.* Nanoscale NMR spectroscopy and imaging of multiple nuclear species. *Nat. Nanotechnol.* **10**, 129–134 (2015).
4. Lovchinsky, I. *et al.* Nuclear magnetic resonance detection and spectroscopy of single proteins using quantum logic. *Science* **351**, 836–841 (2016).
5. Aslam, N. *et al.* Nanoscale nuclear magnetic resonance with chemical resolution. *Science* **357**, 67–71 (2017).
6. Berger, S & Sicker, D. *Classics in Spectroscopy Isolation and Structure Elucidation of Natural Products* 1st edn (Wiley-VCH, 2009).

7. Walsworth, R. L., Glenn, D. R. & Bucher, D. B. Synchronized-readout for narrowband detection of time-varying electromagnetic fields using solid state spins. US patent application 62/341,497 (filed 25 May 2016); International patent application PCT/US17/34256 (filed 24 May 2017).
8. Schmitt, S. *et al.* Submillihertz magnetic spectroscopy performed with a nanoscale quantum sensor. *Science* **356**, 832–837 (2017).
9. Boss, J. M., Cuijia, K. S., Zopes, J. & Degen, C. Quantum sensing with arbitrary frequency resolution. *Science* **356**, 837–840 (2017).
10. Schirhagl, R. *et al.* Nitrogen-vacancy centers in diamond: nanoscale sensors for physics and biology. *Annu. Rev. Phys. Chem.* **65**, 83–105 (2014).
11. Herzog, B. E., Cadeddu, D., Xue, F., Peddibhotla, P. & Poggio, M. Boundary between the thermal and statistical polarization regimes in a nuclear spin ensemble. *Appl. Phys. Lett.* **105**, 043112 (2014).
12. Meriles, C. A. *et al.* Imaging mesoscopic nuclear spin noise with a diamond magnetometer. *J. Chem. Phys.* **133**, 124105 (2010).
13. Staudacher, T. *et al.* Probing molecular dynamics at the nanoscale via an individual paramagnetic centre. *Nat. Commun.* **6**, 8527 (2015).
14. Zaiser, S. *et al.* Enhancing quantum sensing sensitivity by a quantum memory. *Nat. Commun.* **7**, 12279 (2016).
15. Roskopf, T., Zopes, J., Boss, J. M. & Degen, C. L. A quantum spectrum analyzer enhanced by a nuclear spin memory. *npj Quantum Inf.* **3**, 33 (2017).
16. Bloembergen, M., Purcell, E. M. & Pound, R. V. Relaxation effects in nuclear magnetic resonance absorption. *Phys. Rev.* **73**, 679–712 (1948).
17. Olson, D. L., Peck, T. L., Webb, A. G., Magin, R. L. & Sweedler, J. V. High-resolution microcoil ¹H-NMR for mass-limited, nanoliter-volume samples. *Science* **270**, 1967–1970 (1995).
18. Fratila, R. M. & Velders, A. H. Small-volume nuclear magnetic resonance spectroscopy. *Annu. Rev. Anal. Chem.* **4**, 227–249 (2011).
19. Ryan, H., Smith, A. & Utz, M. Structural shimming for high-resolution nuclear magnetic resonance spectroscopy in lab-on-a-chip devices. *Lab Chip* **14**, 1678–1685 (2014).
20. Liao, S. H. *et al.* A study of J-coupling spectroscopy using the Earth's field nuclear magnetic resonance inside a laboratory. *Rev. Sci. Instrum.* **81**, 104104 (2010).
21. Babij, N. R. *et al.* NMR chemical shifts of trace impurities: industrially preferred solvents used in process and green chemistry. *Org. Process Res. Dev.* **20**, 661–667 (2016).
22. Waddington, D. E. J. *et al.* Nanodiamond-enhanced MRI via *in situ* hyperpolarization. *Nat. Commun.* **8**, 15118 (2017).
23. Abrams, D., Trusheim, M. E., Englund, D. R., Shattuck, M. D. & Meriles, C. A. Dynamic nuclear spin polarization of liquids and gases in contact with nanostructured diamond. *Nano Lett.* **14**, 2471–2478 (2014).
24. Mousoulis, C., Xu, X., Reiter, D. A. & Neu, C. P. Single cell spectroscopy: noninvasive measures of small-scale structure and function. *Methods* **64**, 119–128 (2013).
25. Grisi, M. *et al.* NMR spectroscopy of single sub-nL ova with inductive ultra-compact single-chip probes. *Sci. Rep.* **7**, 44670 (2017).
26. Grant, S. C. *et al.* NMR spectroscopy of single neurons. *Magn. Reson. Med.* **44**, 19–22 (2000).
27. Fessenden, M. Metabolomics: Small molecules, single cells. *Nature* **540**, 153–155 (2016).
28. Marusyk, A., Almendro, V. & Polyak, K. Intra-tumour heterogeneity: a looking glass for cancer? *Nat. Rev. Cancer* **12**, 323–334 (2012).
29. Glenn, D. R. *et al.* Single-cell magnetic imaging using a quantum diamond microscope. *Nat. Methods* **12**, 736–738 (2015).
30. Teschler, J. K. *et al.* Living in the matrix: assembly and control of *Vibrio cholerae* biofilms. *Nat. Rev. Microbiol.* **13**, 255–268 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements This material is based on work supported by, or supported in part by, the US Army Research Laboratory and the US Army Research Office under contract/grant number W911NF1510548. D.B.B. was partially supported by the German Research Foundation (BU 3257/1-1). M.D.L. acknowledges support from the Gordon and Betty Moore foundation. We thank R. Fu for assisting with acquisition of the electromagnet used to apply the bias field, M. Rosen for guidance on NMR techniques and S. DeVience for assisting with SpinDynamica calculations of NMR spectra at low B_0 .

Author Contributions R.L.W., D.R.G. and D.B.B. devised the synchronized readout technique. D.R.G., D.B.B. and R.L.W. conceived the application of the technique to picolitre-scale NMR measurements, designed the experiments and analysed the data. D.B.B. and D.R.G. designed and constructed the ensemble NV magnetometer and wrote the accompanying control software. J.L. carried out the single-NV synchronized readout experiments with incoherent averaging. M.D.L., H.P. and R.L.W. conceived the application of NV diamond magnetometry to NMR detection at short length scales. All authors discussed the results and participated in writing the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to R.L.W. (rwalsworth@cfa.harvard.edu).

Reviewer Information Nature thanks D. Budker and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

NV ensemble sensor. The NV ensemble sensor was based on a 2 mm × 2 mm × 0.5 mm diamond chip, created using high-purity chemical vapour deposition (CVD) with 99.999% ^{12}C isotopic purity and a bulk nitrogen concentration of $[^{14}\text{N}] < 8.5 \times 10^{14} \text{ cm}^{-3}$ (Element Six). Modification of the CVD gas mix during the final stage of growth yielded a 13- μm -thick nitrogen-enriched top layer ($[^{14}\text{N}] \approx 4.8 \times 10^{18} \text{ cm}^{-3}$, measured by secondary ion mass spectrometry). The diamond was electron-irradiated (flux of $1.3 \times 10^{14} \text{ cm}^{-2} \text{ s}^{-1}$) for 5 h, and annealed in vacuum (800 °C) for 12 h, yielding an NV concentration of $[\text{NV}] \approx 3 \times 10^{17} \text{ cm}^{-3}$. The ensemble T_2^* dephasing time for NV centres in this diamond, measured using Ramsey spectroscopy, was $T_2^* \approx 750 \text{ ns}$. The NV-ensemble T_2 decoherence time, measured using a Hahn-echo sequence, was 6.5 μs .

For NMR measurements, the diamond was cut so that the top face was perpendicular to the [100] crystal axis and the lateral faces were perpendicular to [110]. All four edges of the top face were then polished at 45° (Delaware Diamond Knives), resulting in a truncated square pyramid, with a top face area of 1 mm × 1 mm. This angle polishing allowed total internal reflection of the laser beam (see Fig. 1a) to prevent direct illumination of the NMR sample. Excitation light was provided by a diode-pumped solid-state laser at 532 nm (Coherent Verdi G7), directed through an acousto-optic modulator (AOM) (IntraAction ASM802B47) to produce 5- μs pulses. The first approximately 1 μs of each pulse was used to optically read the spin state of the NV ensemble and the remainder of the pulse repolarized the NVs. The AOM was driven by a digitally synthesized 80-MHz sinusoid (Tektronix AWG 7122C), amplified to 33 dBm (Minicircuits ZHL-03-5WF), and the total laser power at the sensor volume was 150 mW. The laser was focused to an approximately 20- μm -diameter waist near the position of the NV sensor layer, resulting in an optical intensity of 48 kW cm $^{-2}$ (comparable to the typical NV saturation intensity $I_{\text{sat}} \approx 100 \text{ kW cm}^{-2}$).

For all NMR experiments, the diamond was mounted by gluing (Epoxy Technology Inc., EPO-TEK 301) to a 3-mm glass prism (Thorlabs PS905) and placed inside a sample cuvette (FireflySci Type4 Microfluorescence Cuvette). The diamond was then rotated so that a [111] diamond crystal axis was aligned to the static bias magnetic field B_0 . NV centres aligned along this axis were used for sensing, whereas those along the other three [111] directions were far off-resonance and contributed only to the background fluorescence. (The background fluorescence contribution from non-aligned NV centres was minimized by adjusting the linear polarization angle of the excitation laser.) The magnetic-field alignment was carried out by overlapping the pulsed electron spin resonance (ESR) frequencies of the three non-aligned axes with one another. The bias magnetic field strength was $B_0 = 88 \text{ mT}$, such that the resonance frequency of the $|m_s = 0\rangle \rightarrow |m_s = -1\rangle$ spin transition for the aligned NV centres was $f_{\text{NV-Larmor}} = 400 \text{ MHz}$. (The $|m_s = 0\rangle \rightarrow |m_s = +1\rangle$ resonance frequency was 5,340 MHz.)

NV magnetometry pulse sequences for magnetic resonance detection were carried out on the $|m_s = 0\rangle \rightarrow |m_s = -1\rangle$ transition. Microwaves were delivered using a straight length of wire (0.25-mm diameter) positioned above the diamond, approximately 0.4 mm away from the NV sensing volume. The 400-MHz carrier frequency and the pulse modulation were both synthesized digitally (Tektronix AWG 7122C); pulses were then amplified to 40 dBm (Minicircuits ZHL-100W-52-S+) and coupled into the wire, yielding a NV Rabi frequency of $f_1 = 5.6 \text{ MHz}$. (The maximum achievable NV Rabi frequency using the full output of the amplifier was $f_1 = 16.6 \text{ MHz}$.) An XY8-4 or XY8-6 dynamical decoupling sequence was used to detect magnetic resonance signals selectively around 3.755 MHz, which is the proton Larmor frequency at $B_0 = 88 \text{ mT}$. The phase of the final $\pi/2$ pulse of the sequence was optimized to give a fluorescence corresponding to a mixed state of the NV (that is, equal to the mean fluorescence over one Rabi oscillation), to make the fluorescence signal linearly sensitive to small magnetic-field amplitudes. For an ideal two-level quantum system, this condition would correspond to a 90° phase shift between the initial and final $\pi/2$ pulses; in practice, the small drive detunings associated with the ^{14}N hyperfine structure required manual optimization of the phase. To reject laser intensity noise and microwave power fluctuations, the phase of the final $\pi/2$ pulse of every second synchronized readout magnetometry subsequence was shifted by 180° relative to the nominal value, and successive pairs of readouts were amplitude-subtracted. Thus, one synchronized readout time-series data point was recorded for every two magnetometry subsequences.

Spin-state-dependent fluorescence from the NV centres was collected with a glass (BK-7) light guide (Edmund Optics 5-mm Aperture, 120 mm L, Low NA Hexagonal Light Pipe) and delivered to a balanced photodiode module (Thorlabs PDB210A). To eliminate scatter from the excitation laser, an interference filter (Semrock BLP01-647R) was placed between the light guide and the detector. A small fraction of the excitation beam was split off upstream of the diamond chip and directed to the second channel of the balanced diode module. A glass slide mounted on a motorized stage (Thorlabs PRM1Z8) in the second path enabled

automated re-balancing between averages during long synchronized readout signal acquisitions. When the NV centres were fully polarized in $|m_s = 0\rangle$, the light-induced fluorescence signal produced a single-channel (unbalanced) photocurrent of 30 μA . Immediately after applying a microwave π pulse, the single-channel photocurrent was 28 μA , indicating a maximum fluorescence contrast of about 7%. The difference signal of the photodiode module (with onboard transimpedance gain of $1.75 \times 10^5 \text{ V A}^{-1}$) was further amplified by 3 dB and low-pass-filtered at 1 MHz using a low-noise pre-amplifier unit (Stanford Research SR-560), then recorded with a digital-to-analogue converter (DAQ) (National Instruments NI-USB 6281). The DAQ bandwidth was 750 kHz and the digitization was on-demand, triggered by a transistor-transistor logic (TTL) pulse from the arbitrary-waveform generator (AWG) used to control the experiment. The delay between the rising edges of the AOM gate pulse and the DAQ trigger was optimized for maximum spin-state-dependent fluorescence signal.

Synchronized readout protocol and data analysis. The synchronized readout protocol, including CASR detection of NMR, consists of a series of repeated NV AC magnetometry subsequences, with each subsequence followed by a projective NV spin-state readout. The protocol is defined with respect to a particular central frequency f_0 . The repeated AC magnetometry subsequences are all identical, each consisting of an initial $\pi/2$ pulse, followed by a train of π pulses applied at a rate of $2f_0$, and ending with another $\pi/2$ pulse. The initial and final $\pi/2$ pulses are chosen to have a relative phase shift, such that the final NV spin population is linearly dependent on the amplitude of the oscillating magnetic-field signal. (The choice of linear, rather than quadratic, dependence of the final NV population on magnetic field strength is important for minimizing sensitivity to noise due to sample fluctuations and diffusion; see Supplementary Note 10). At every synchronized readout step, the accumulated NV spin population is measured via spin-state-dependent fluorescence and the NV spin is optically repolarized. The delay between the start of successive magnetometry subsequences in the synchronized readout protocol is an integer number k of periods at the central frequency, $\tau_{\text{SR}} = k/f_0$. The range of signal frequencies f that can be detected without aliasing is therefore given by $f_0 \pm 1/(2\tau_{\text{SR}})$.

In the experiments described here, the synchronized readout cycle period τ_{SR} , the reciprocal central synchronized readout detection frequency $1/f_0 = 1/(3.74065 \text{ MHz}) = 267.3 \text{ ns}$, and the reciprocal NV drive frequency $1/f_{\text{NV-Larmor}} = 1/(400 \text{ MHz}) = 2.5 \text{ ns}$ were all chosen to be exact integer multiples of the clock period of the timing generator (Tektronix AWG 7122C), $\tau_{\text{clock}} = 1/(12 \text{ GHz}) = 0.083 \text{ ns}$. (The use of integer frequency sub-multiples effectively determined the exact value of B_0 used in our experiments; non-integer sub-multiples may be used if necessary, at the expense of introducing a series of discrete spectral artefacts into the synchronized readout spectrum.) The NV magnetometry pulse sequence (XY8-4 or XY8-6) was saved in the memory of the AWG and its output was gated by a TTL signal from a programmable pulse generator (Spincore PulseBlasterESR-PRO 500 MHz). The PulseBlaster gate determined the number of synchronized readout iterations per experiment n_{SR} , resulting in a synchronized readout measurement duration of $T = n_{\text{SR}}\tau_{\text{SR}}$. When detecting the NMR signals using CASR (in experiments shown in Figs 2, 3), the pulse blaster also generated the TTL pulse for gating the proton-driving radio-frequency pulses. All FNP experiments used a single proton $\pi/2$ pulse at the start of the experiment ($t = 0$); the water echo experiment (Fig. 2c) used proton π pulses also at $t = 40 \text{ ms}$ and $t = 120 \text{ ms}$. Each readout of the synchronized readout protocol was saved in a numerical array, giving a time series of length n_{SR} . Individual time series were averaged N_{avg} times to improve the signal-to-noise ratio. A summary of the synchronized readout parameters for each experiment described in the main text is as follows: glycerol FNP (Fig. 2a), XY8-4 sub-sequences, $\tau_{\text{SR}} = 24.06 \mu\text{s}$, $n_{\text{SR}} = 4 \times 10^4$, $T = 0.962 \text{ s}$, $N_{\text{avg}} = 7.5 \times 10^4$; glycerol FNP (Fig. 2b), XY8-4 sub-sequences, $\tau_{\text{SR}} = 24.06 \mu\text{s}$, $n_{\text{SR}} = 2 \times 10^3$, $T = 0.048 \text{ s}$, $N_{\text{avg}} = 5.8 \times 10^4$; glycerol FNP (Fig. 2c), XY8-4 sub-sequences, $\tau_{\text{SR}} = 24.06 \mu\text{s}$, $n_{\text{SR}} = 4 \times 10^4$, $T = 0.962 \text{ s}$, $N_{\text{avg}} = 7.5 \times 10^4$; water FNP (Fig. 2c), XY8-4 sub-sequences, $\tau_{\text{SR}} = 24.06 \mu\text{s}$, $n_{\text{SR}} = 8 \times 10^4$, $T = 1.925 \text{ s}$, $N_{\text{avg}} = 1.6 \times 10^4$; water echo (Fig. 2c), XY8-4 sub-sequences, $\tau_{\text{SR}} = 24.06 \mu\text{s}$, $n_{\text{SR}} = 8 \times 10^4$, $T = 1.925 \text{ s}$, $N_{\text{avg}} = 2.0 \times 10^4$; molecule FNPs (Fig. 3a, b), XY8-6 sub-sequences, $\tau_{\text{SR}} = 24.06 \mu\text{s}$, $n_{\text{SR}} = 4 \times 10^4$, $T = 0.962 \text{ s}$, $N_{\text{avg}} = 3.4 \times 10^4$; molecule FNP (Fig. 3c), XY8-6 sub-sequences, $\tau_{\text{SR}} = 24.06 \mu\text{s}$, $n_{\text{SR}} = 4 \times 10^4$, $T = 0.962 \text{ s}$, $N_{\text{avg}} = 3.6 \times 10^4$; coil signal (Extended Data Fig. 6c), XY8-4 sub-sequences, $\tau_{\text{SR}} = 1.2 \text{ ms}$, $n_{\text{SR}} = 2.5 \times 10^6$, $T = 3,000 \text{ s}$, $N_{\text{avg}} = 1$.

For CASR-NMR measurements, the first 20 time-series data points, which coincided with the proton $\pi/2$ pulse plus approximately 50 times the coil ring-down time, were discarded from the recorded time-series data. After averaging in the time domain, the data were mean-subtracted and then Fourier transformed and fitted using MATLAB. Each spectrum was fitted to both Lorentzian and Gaussian line shapes, and the model with smaller residuals (the Lorentzian in all cases except that of Fig. 3b) was selected for display. Unless otherwise speci-

fied, all spectra shown in the figures are power spectra, calculated as the squared absolute value of the Fourier-transformed time-series data. (This differs from the convention of standard NMR, where spectra are typically presented as the real or imaginary part of the Fourier transform; the difference is accounted for in the form of the fit functions used.) Before squaring to obtain the power spectrum, the complex Fourier-transformed data were smoothed (boxcar convolution) using a square filter width of 1.0 Hz (Fig. 2c, spin echo), 1.5 Hz (Fig. 2c, FNPs), 2.5 Hz (Fig. 3a, b) or 3.0 Hz (Fig. 3c). When uncertainties are quoted for the spectral line width or the splitting parameters, they were estimated by repeating the full experiment (including all averages) and fitting procedure $n \geq 3$ times, then calculating the sample standard deviation

$$\sigma = \left[\sum_j \frac{(x_j - \langle x \rangle)^2}{n-1} \right]^{1/2}$$

over the ensemble of fitted parameters. Uncertainties reported are 1σ in every case.

Electromagnet. The bias magnetic field B_0 was produced by an air-cooled electromagnet (Newport Instruments Type A). The pole pieces were cylindrical, 10 cm in diameter, with an adjustable gap set to 3 cm. The main coils (each 1,900 turns of copper strip, with room-temperature resistance $R = 4.5 \Omega$) were driven (Hewlett Packard HP 6274) with a continuous current of about 650 mA to produce a nominal field of $B_0 \approx 88$ mT. We operated at $B_0 \leq 0.1$ T to obtain a relatively low NV Larmor frequency $f_{\text{NV-Larmor}} = 400$ MHz, at which inexpensive power amplifiers for generating NV drive pulses are readily available. The choice of $B_0 = 88$ mT enabled the NV and proton Larmor frequencies to be integer sub-multiples of the AWG clock frequency, as described in Methods section ‘Synchronized readout protocol and data analysis’. A secondary coil pair (diameter, 10 cm; gap, 7 cm; 15 turns each) was manually wound around the poles to enable precise field stabilization without the need for very small adjustments to the main current supply. The secondary coils were driven by a voltage-controlled current supply (Thorlabs LDC205C), controlled by the analogue output channel of a DAQ (National Instruments PCI 6036E).

Magnetic bias field stabilization. The static bias field B_0 generated by the electromagnet was stabilized actively using two feedback systems. On short timescales, the field stabilization relied on a secondary NV-diamond magnetometer performing continuous-wave ESR measurements. The secondary magnetometer was positioned between the electromagnet poles, approximately 1 cm away from the primary ensemble NV-NMR sensor (Extended Data Fig. 2). The continuous-wave ESR microwave-frequency modulation was locked to the main synchronized readout experiment using the same AWG (Tektronix AWG 7122C) to ensure that any cross-talk between the detectors was coherent over averages of the synchronized readout protocol and could be removed during data analysis. (This precaution proved unnecessary in most of the CASR NMR experiments because the continuous-wave ESR drive power was too weak to produce a measureable effect on the CASR sensor.) The excitation laser, light collection optics and microwave drive for the secondary experiment were all independent from those of the main synchronized readout magnetic resonance sensor. This enabled feedback control over magnetic-field fluctuations (primarily due to current noise in the main coils) with a bandwidth of about 12.5 Hz, resulting in short-term (around 30 min) field stability, $\sigma_B \approx 15$ nT (root-mean-square).

To correct slow drifts between the main magnetic resonance sensor and the secondary field-stabilization sensor, we paused the synchronized readout protocol between averages periodically (every 5 min) and performed pulsed ESR

measurements on the primary NV-diamond sensor. Any measured magnetic-field drifts were used to correct the set-point of the fast feedback loop to ensure long-term (about 50 h) stability, $\sigma_B \approx 23$ nT (root-mean-square). We therefore conclude that residual B_0 fluctuations limit the observed proton NMR line width to approximately $\Gamma \approx 2[2\ln(2)]^{1/2} \times 23 \text{ nT} \times 42.58 \text{ MHz T}^{-1} = 2.3 \text{ Hz}$ (FWHM).

All continuous-wave ESR measurements were carried out using both the $|m_s = 0\rangle \rightarrow |m_s = -1\rangle$ and the $|m_s = 0\rangle \rightarrow |m_s = +1\rangle$ transitions of the aligned NV centres, to distinguish resonance shifts due to changes in temperature³¹ and magnetic field. For fast-feedback measurements on the secondary field-stabilization sensor, we monitored only four discrete optically detected magnetic resonance (ODMR) frequencies to maximize bandwidth. This system was potentially susceptible to second-order feedback errors associated with simultaneous changes in B_0 and temperature. We therefore anchored the secondary field-stabilization sensor thermally to a piece of black-anodized aluminium and stabilized its temperature actively using absorption from a separate DPSS laser (Thorlabs DJ532-40). Temperature control was not required for slow feedback on the main magnetic resonance sensor, where we acquired a full ODMR spectrum (58 frequency points) to fully account for all drifts in magnetic field, temperature and optical contrast.

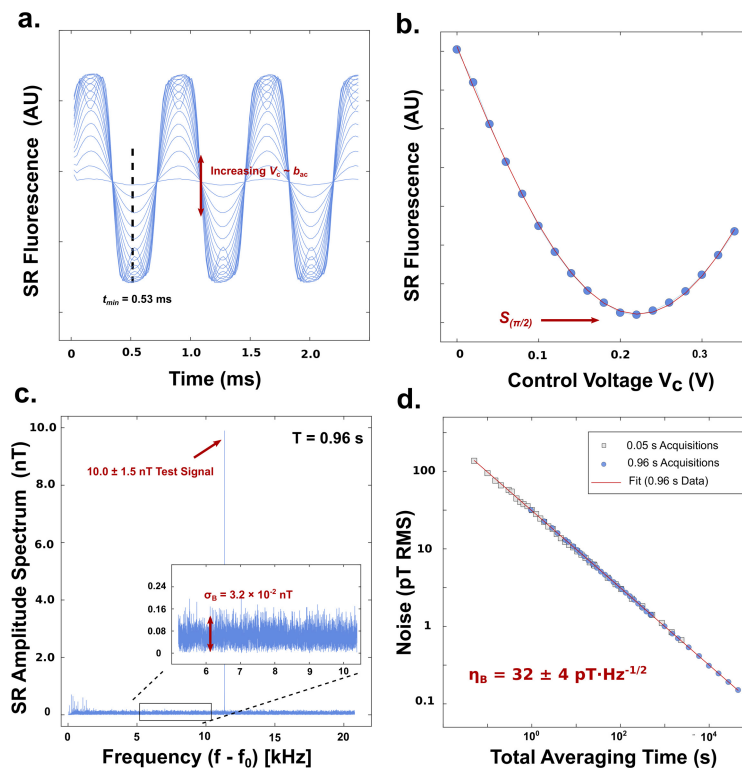
NMR drive coils. Radio-frequency pulses for driving sample proton spins were produced by a pair of cylindrical coils wound around the sample cuvette. This geometry, with 1.1-cm coil diameter and 1.2-cm centre-to-centre spacing, provided a combination of strong drive fields and convenient optical access to the NV-ensemble sensor. The coils were 22 turns each, connected in series and coupled to the current source (Rigol DG 1032) with a standard network of variable matching and coupling capacitors³². After tuning, the resonance frequency was 3.75 MHz and the quality factor (Q) of the coil was 140. Driving the coils on resonance, we obtained a maximum nuclear-spin Rabi frequency of $f_{\Omega, \text{nuc}} \approx 8$ kHz.

NMR samples. Deionized water was obtained from Ricca Chemical Company (part number 9150-5). *p*-xylene, glycerol, trimethyl phosphate and ethyl formate were purchased from Sigma Aldrich (catalogue numbers 296333, G9012, 241024 and 112682, respectively) and used without dilution or modification. The glycerol sample may have contained some atmospherically absorbed water (less than 20% by volume).

Comparison between CASR and microcoils. We compared the sensitivity and measurement volume of NV-NMR sensing using CASR with recently demonstrated micrometre-scale inductive detectors (Extended Data Fig. 8). The comparison was based primarily on data from a review of microscale detectors³³, with the addition of two more recent results^{25,34}. We excluded data points from the review that correspond to demonstrations of magnetic resonance imaging; only measurements of multi-component spectra were included. Sensitivities for the inductive detectors were extrapolated to a common bias field of $B_0 = 14.1$ T.

Data availability. The data that support the findings of this study are available from the corresponding author on reasonable request. Source Data for Figs 2 and 3 are available with the online version of the paper.

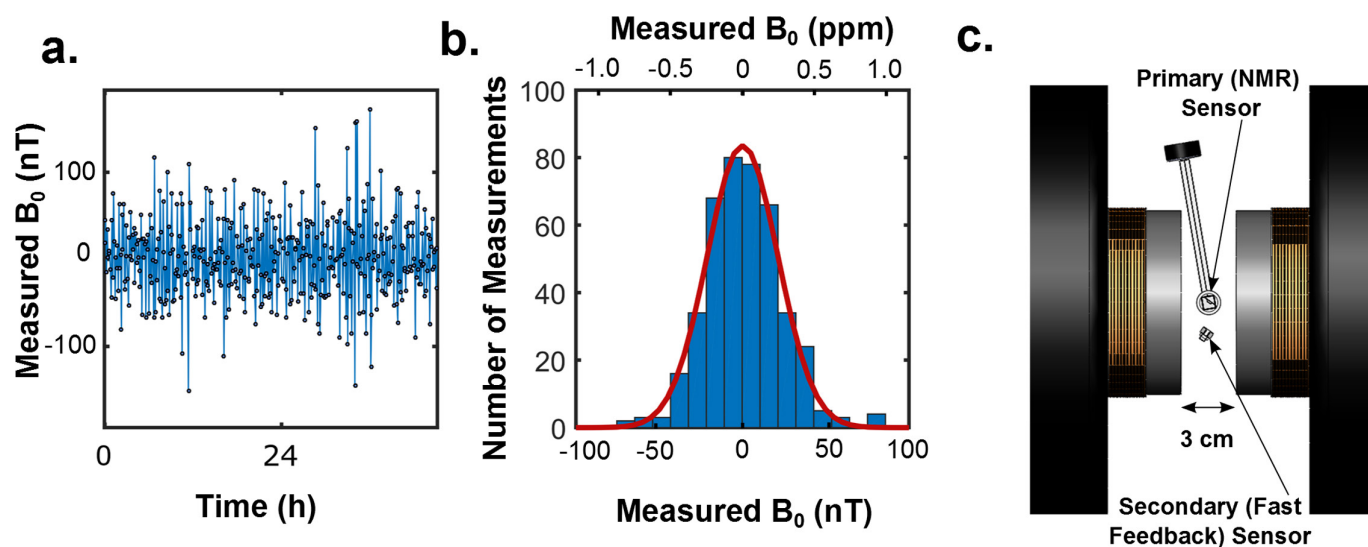
- Acosta, V. M. *et al.* Temperature dependence of the nitrogen-vacancy magnetic resonance in diamond. *Phys. Rev. Lett.* **104**, 070801 (2010).
- Wheeler, D. D. & Conradi, M. S. Practical exercises for learning to construct NMR/MRI probe circuits. *Concepts Magn. Reson.* **40A**, 1–13 (2012).
- Badilita, V. *et al.* Microscale nuclear magnetic resonance: a tool for soft matter research. *Soft Matter* **8**, 10583–10597 (2012).
- Finch, G., Yilmaz, A. & Utz, M. An optimized detector for *in-situ* high-resolution NMR in microfluidic devices. *J. Magn. Reson.* **262**, 73–80 (2016).



Extended Data Figure 1 | Sensitivity of the NV-ensemble sensor.

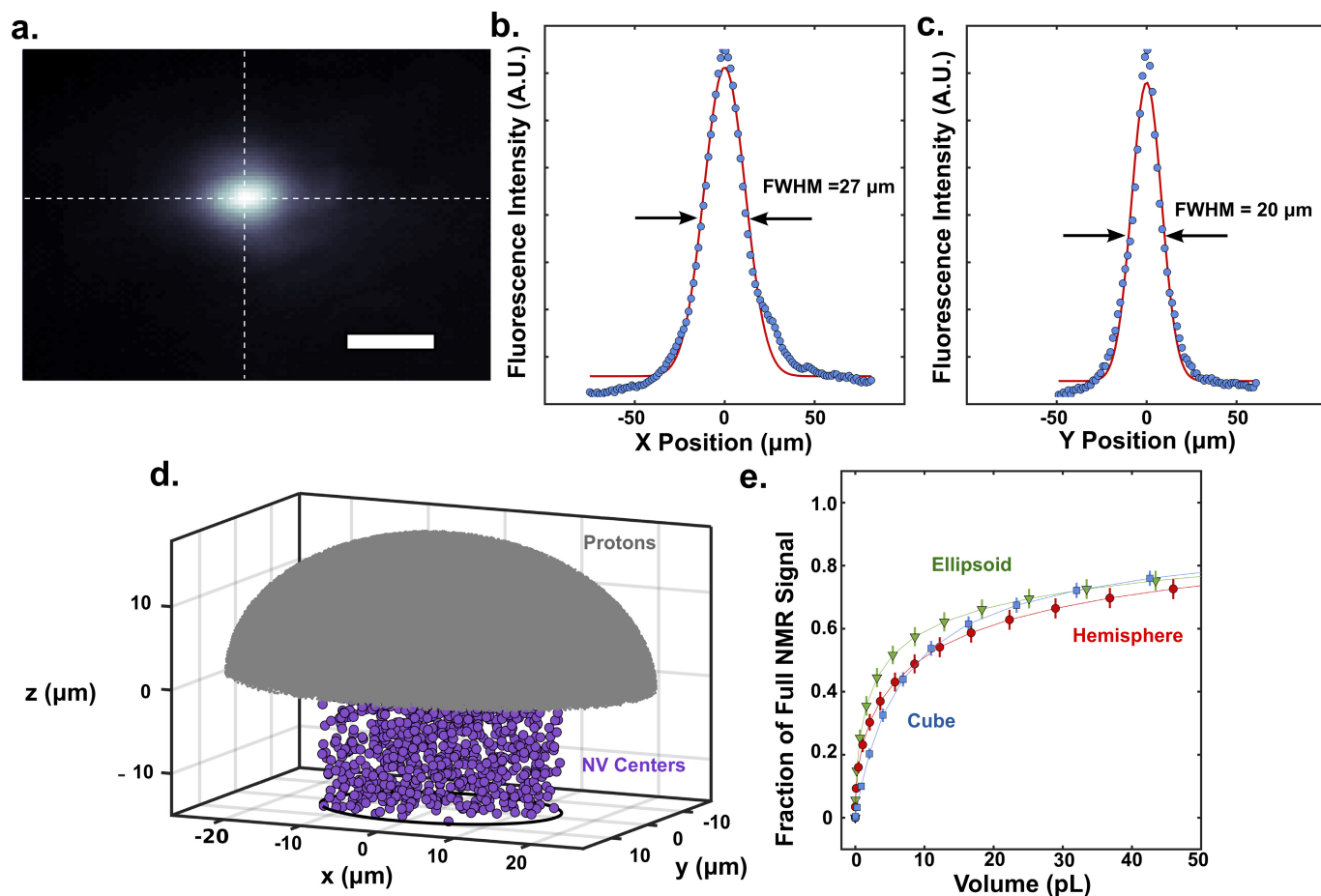
a, Synchronized readout (SR) measurements of magnetic test signals from a nearby coil antenna at $f_{\text{coil}} = 3.742$ MHz (Supplementary Methods 1). The control voltage for the AC current source of the coil was varied from $V_c = 0$ to $V_c = 0.35$ V. Each blue trace corresponds to a different value of the control voltage. The amplitude of the oscillating magnetic field that is produced by the coil (b_{ac}) is proportional to V_c . The amplitude of the measured synchronized readout signal increases with V_c . **b**, Synchronized readout signal data (blue points) as a function of control voltage V_c at constant time $t_{\text{min}} \approx 0.53$ ms, obtained by cutting the data shown in **a** along the dashed line. The red line is a sinusoidal fit to the data, from which we obtain the control voltage $V_c = 0.22$ V that produces a $\pi/2$ NV phase accumulation in a single magnetometry subsequence. The $\pi/2$ NV phase

accumulation occurs when the fluorescence signal is at its minimum, $S_{(\pi/2)}$. This provides a calibration for the amplitude of the applied test signal. **c**, Synchronized readout amplitude spectrum of a 10.0-nT test signal ($f_{\text{coil}} = 3.752$ MHz), recorded in $T = 0.96$ s. The calibrated signal amplitude defines the vertical axis of the plot. The rectangular window shows the frequency range used to estimate the noise in the spectrum. The noise amplitude σ_B is determined by comparing with the calibrated test signal. **d**, Synchronized readout noise measurements as a function of averaging time for acquisition durations of 0.96 s (blue circles) and 0.05 s (grey boxes). A power-law fit to the 0.96 s data (red line) indicates an inverse square-root scaling with time and a sensitivity of $\eta_B = 32 \pm 4$ pT Hz $^{-1/2}$. AU, arbitrary units; RMS, root-mean-square.



Extended Data Figure 2 | Electromagnet stabilization. **a**, Time-series data of magnetic field (B_0) deviations, recorded at the primary (NMR) NV-diamond-ensemble sensor, once every 5 min, over 48 h. **b**, Histogram of the data in **a**, showing a Gaussian distribution of magnetic-field deviations with a standard deviation of 46 nT. Because the measurement precision of the sensors and the current precision of the coil used to correct B_0 were both much smaller than the actual B_0 fluctuations, the deviation from the set-point was effectively zero immediately after every

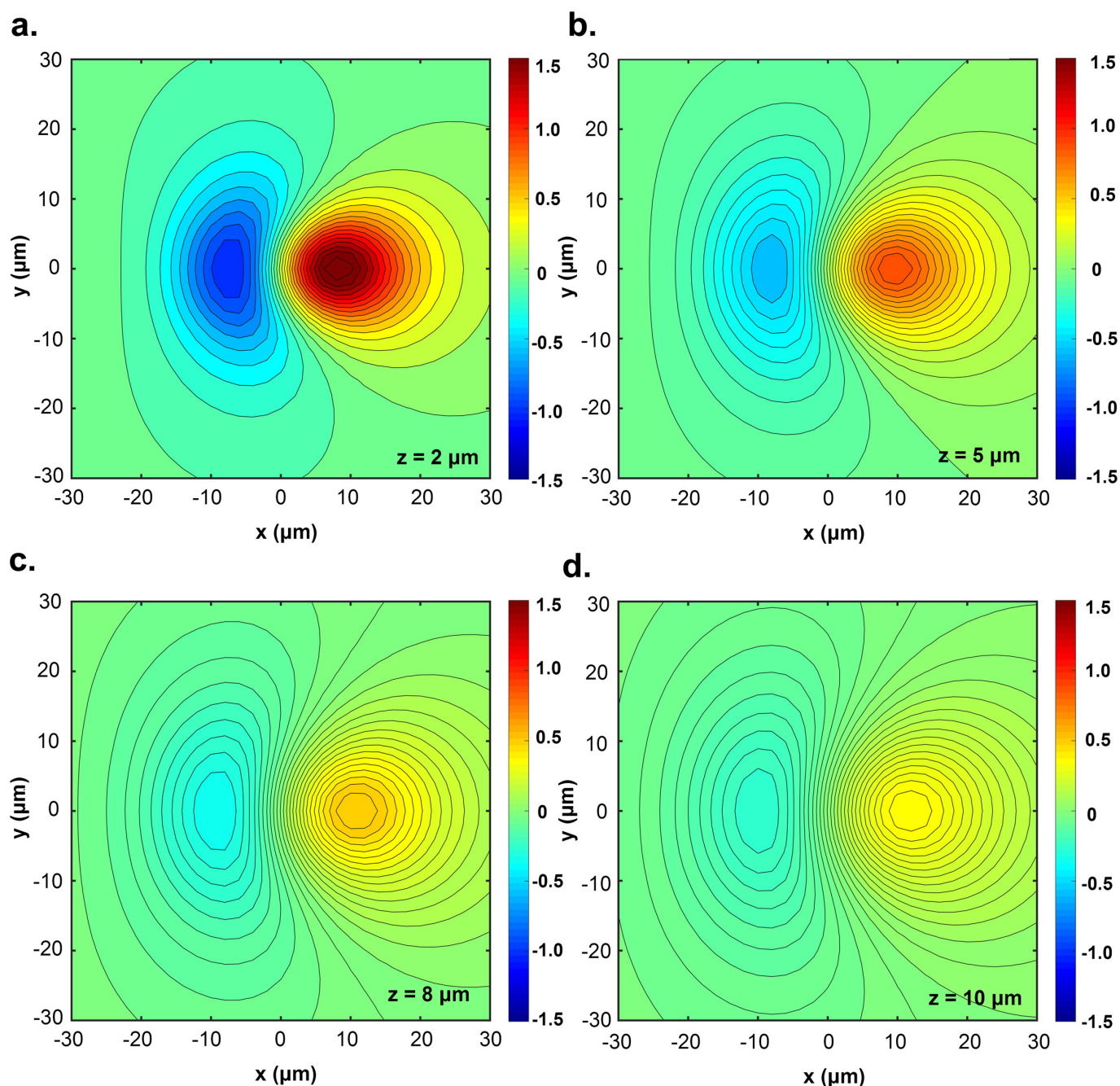
feedback adjustment. Assuming linear drift of the magnetic field during each (5-min) feedback interval, the average magnetic-field deviation from the set-point during the interval was approximately half the value recorded at the end of the interval. The real standard deviation of the magnetic-field fluctuations at the primary NV-NMR sensor was therefore estimated to be about 23 nT. **c**, Schematic of the electromagnet and sensors, drawn to scale. The black coils are the main magnet coils (88 mT); copper coils are the correction coils for fast control of B_0 .



Extended Data Figure 3 | Estimate of NMR measurement volume.

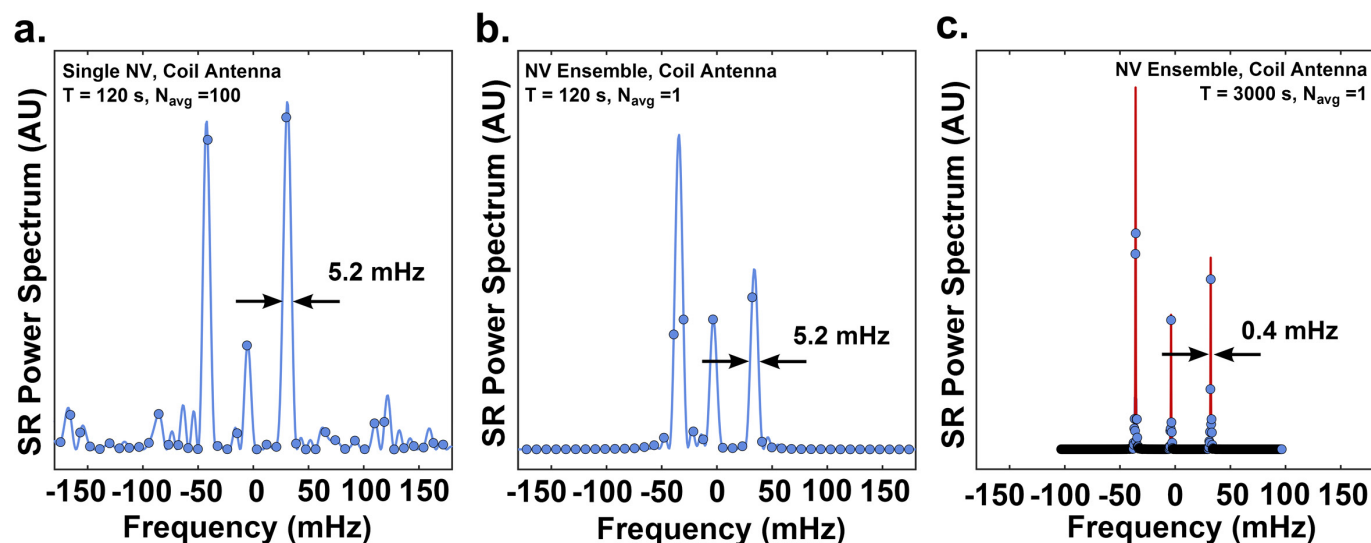
a. Fluorescence image of the NV sensing volume. The image shown has been stretched by $2^{1/2}$ in the horizontal (x) direction to account for the 45° angle between the imaging plane and the diamond surface. Scale bar, $30\text{ }\mu\text{m}$. **b, c.** Cuts through the image in **a** (horizontal, **b**; vertical, **c**) fitted to Gaussian line shapes. The extracted spot size was $27\text{ }\mu\text{m}$ FWHM in x and $20\text{ }\mu\text{m}$ in y . **d.** Example random configuration of NV centres (purple) and sample protons (grey) generated by the Monte Carlo calculation (Supplementary Note 3) used to estimate the total NMR signal magnitude integrated over the sensor. The NV-sensor volume is modelled as an

elliptic cylinder with semi-axes of $14\text{ }\mu\text{m}$ in x and $10\text{ }\mu\text{m}$ in y and a height of $13\text{ }\mu\text{m}$ in z . The set of protons shown corresponds to a 25 pL half-ellipsoid measurement volume. The diamond surface is at $z = 0$. **e.** Calculated NMR signal, integrated over the NV-sensor volume and normalized to the asymptotic NMR signal at large sample volume. The integral over the sample was carried out using half-ellipsoid, hemispheric and cube-shaped volumes; the volume at which the protons contained therein produced a normalized signal of 0.5 was less than 10 pL in each case. Error bars are numerical uncertainties (1σ) obtained from 10 repetitions of the Monte Carlo calculation. A.U., arbitrary units.



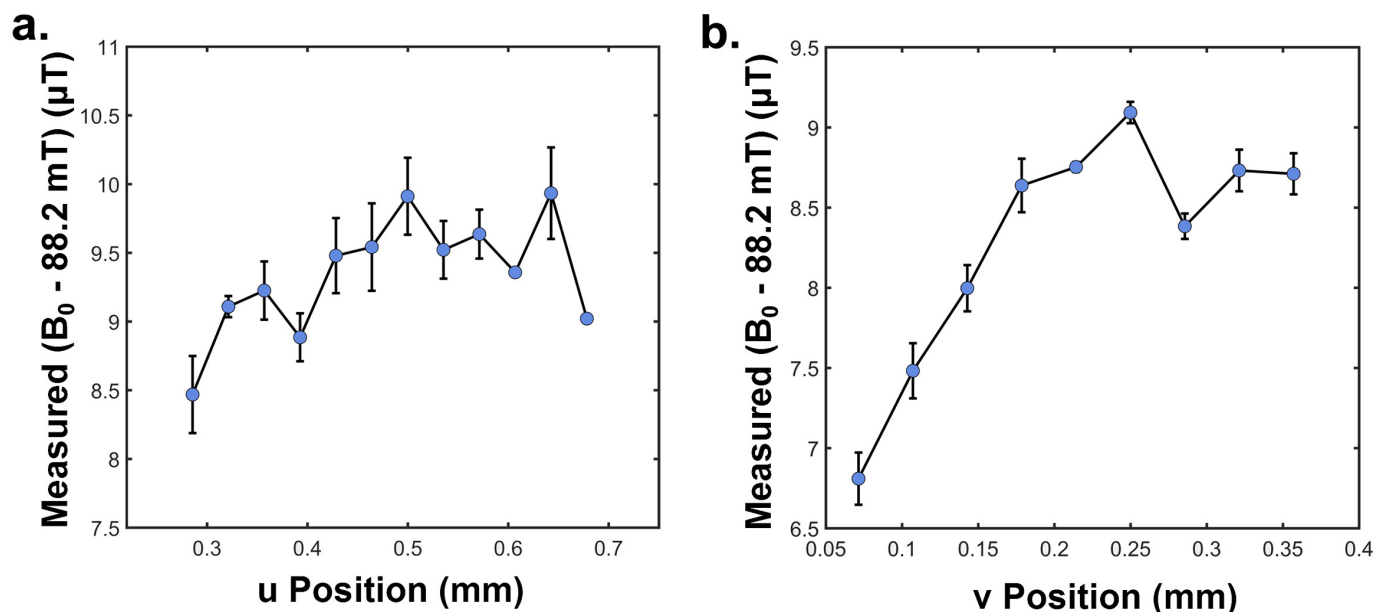
Extended Data Figure 4 | Scaling of the NV magnetic back-action calculated as a function of the position in the NMR sample. a–d, The position-dependent back-action magnetic field in the NMR sample volume, produced by spin-polarized NV centres during CASR sensing, is calculated numerically (Supplementary Note 4). The B -field integral factors into an NV-density-dependent constant (74 nT for the present sensor) and a dimensionless, position-dependent geometric factor (colour scale). The NV magnetization is approximated as a two-dimensional Gaussian in x and y (with FWHMs of $28\mu\text{m}$ and $20\mu\text{m}$, respectively) to represent the laser-intensity-dependent NV polarization, and as a step function in z (so that it is non-zero only between $z = -13\mu\text{m}$ and $0\mu\text{m}$)

to represent the finite extent of the NV layer below the diamond surface. The geometric factor is calculated in several planes above the diamond surface, including $z = 2\mu\text{m}$ (a), $z = 5\mu\text{m}$ (b), $z = 8\mu\text{m}$ (c) and $z = 10\mu\text{m}$ (d). Even at a distance of only about $2\mu\text{m}$ above the diamond surface, the maximum range of the geometric factor is approximately between -1 and $+1$. This corresponds to an approximately $\pm 74\text{ nT}$ (or about $\pm 1\text{ p.p.m.}$ of B_0) shift in the magnetic field seen by the protons during each CASR magnetometry subsequence due to the NV centres. The back-action shift is much smaller than this magnetic-field shift in most of the measurement volume.



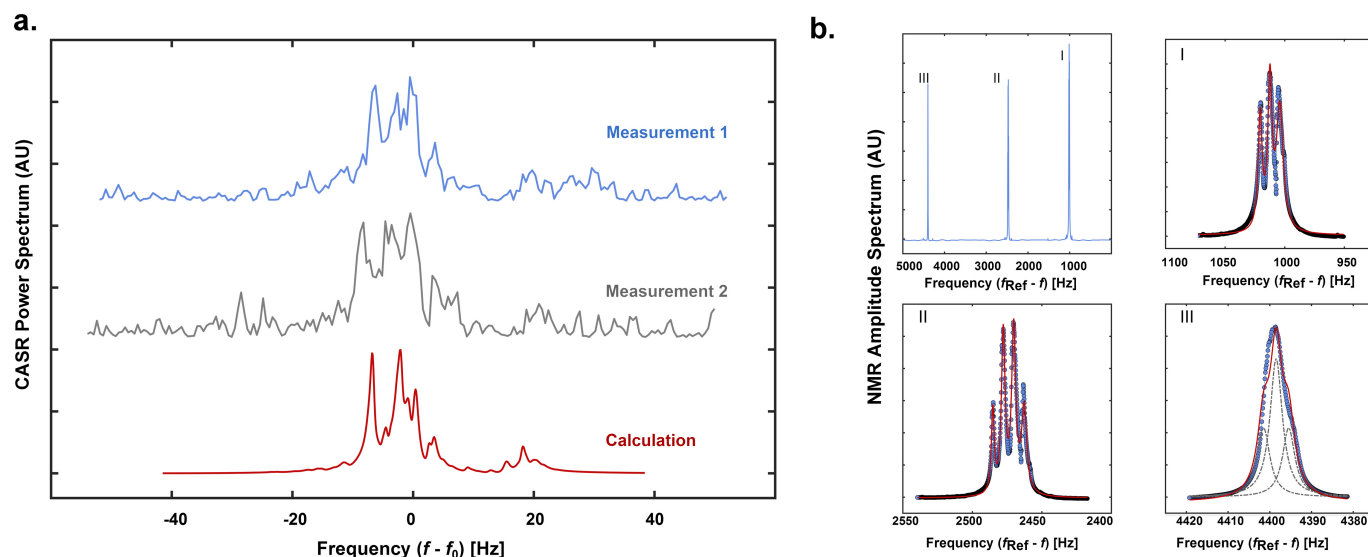
Extended Data Figure 5 | Synchronized readout spectral resolution measured using signals from a coil antenna. **a**, Power spectrum of the synchronized readout signal obtained with a single-NV magnetic sensor in a confocal microscope (Supplementary Methods 5). The synchronized readout protocol used an iteration time of $\tau_{\text{SR}} = 75 \mu\text{s}$ and the total experiment duration was $T = n\tau_{\text{SR}} = 112.5$ s, for $n = 1.5 \times 10^6$ iterations. Data shown are the average of $N_{\text{avg}} = 100$ experiments. The observed spectral width was 5.2 mHz (FWHM). Independent, spectrally narrow signal sources were used to drive each of the three detected frequencies. Successive synchronized readout sequences were incoherently averaged, resulting in poor a signal-to-noise ratio compared to CASR. **b**, Power spectrum of the synchronized readout signal obtained with an NV-ensemble magnetic sensor. The synchronized readout protocol

used an iteration time of $\tau_{\text{SR}} = 75 \mu\text{s}$ and the total experiment duration was $T = n\tau_{\text{SR}} = 112.5$ s, for $n = 1.5 \times 10^6$ iterations. The spectrum shown is for a single average ($N_{\text{avg}} = 1$). The observed spectral width was again 5.2 mHz (FWHM). **c**, Power spectrum of the synchronized readout signal obtained with an NV-ensemble magnetic sensor. The synchronized readout protocol used an iteration time of $\tau_{\text{SR}} = 1.2$ ms and the total experiment duration was $T = n\tau_{\text{SR}} = 3,000$ s, for $n = 2.5 \times 10^6$ iterations. The observed spectral width was 0.4 mHz (FWHM), substantially broader than the Fourier limit. The measured line widths for the three signals were consistent to within about 10%, suggesting that the spectral resolution in this measurement was limited by the stability of the timing source used to control the synchronized readout protocol.



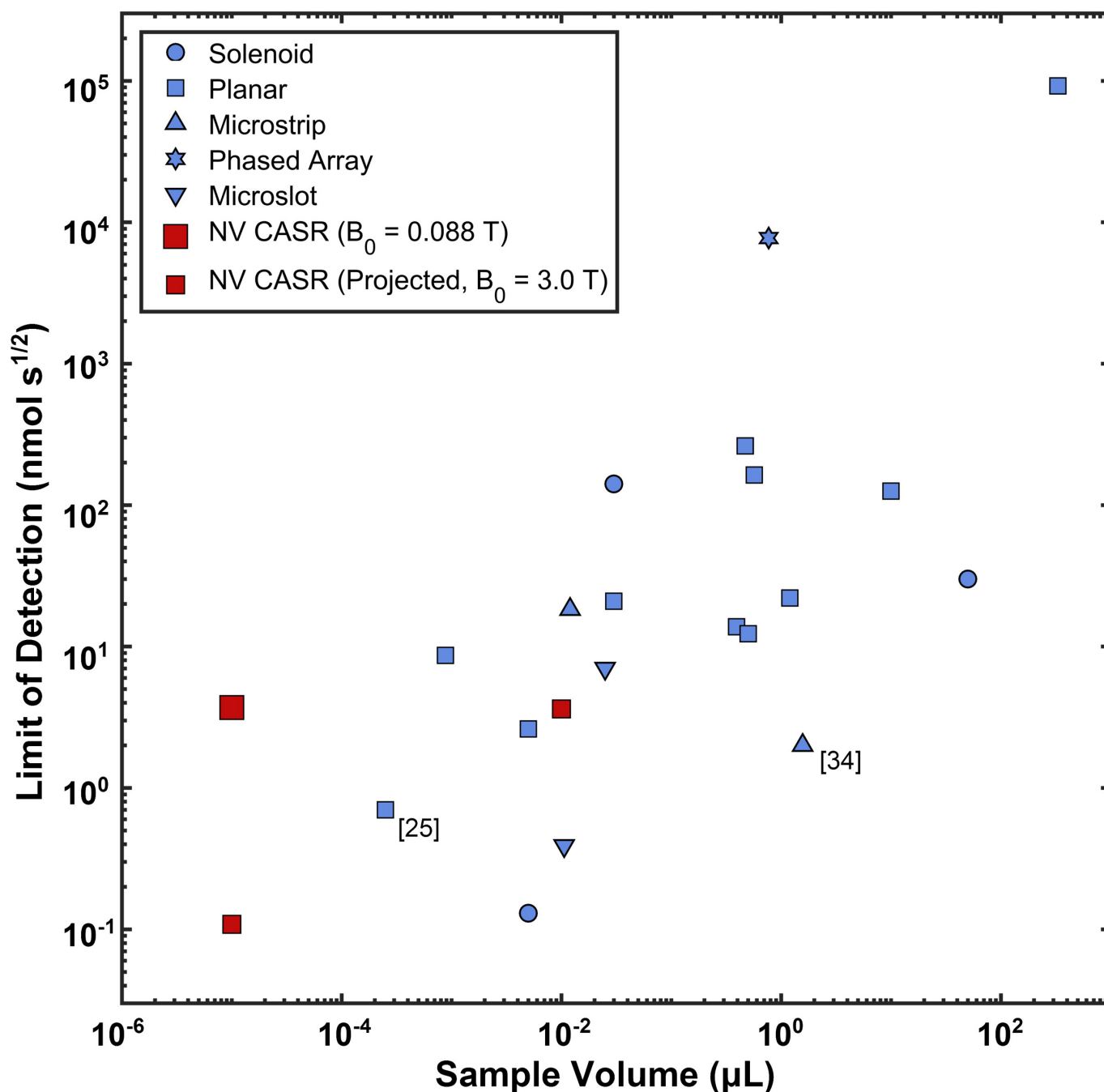
Extended Data Figure 6 | Spatial inhomogeneity of B_0 . **a, b,** Transverse gradients in the magnetic bias field B_0 , sampled in the vicinity of the NMR measurement volume using pulsed ESR applied to the NV centres (Supplementary Methods 6). Local magnetic fields are determined by scanning the excitation laser across the diamond surface in the u direction (parallel to the diamond face, along the line with maximum projection on the cylindrical axis of the magnet poles; **a**) and the v direction (parallel to the diamond face, along the line perpendicular to the magnet axis; **b**). NV-ensemble ESR spectra were recorded at each scan position and the

scan was repeated three times ($n = 3$). Error bars were estimated by computing the standard error in the mean σ/\sqrt{n} of magnetic-field values over the repeated measurements at each scan position. The observed magnetic-field gradient in the v direction, $dB_0/dv \approx 10 \mu\text{T mm}^{-1}$, is expected to yield a FNP spectral signal width of about 8.5 Hz (or about 2.3 p.p.m.) for the NMR measurement volume of diameter about $20 \mu\text{m}$. This value is consistent with the observed FNP line widths of 8–10 Hz for water (Fig. 2c).



Extended Data Figure 7 | Calculated and measured ethyl formate spectra. **a.** Measured and calculated CASR NMR power spectra (offset for clarity) of ethyl formate at $B_0 = 88$ mT. The blue trace is the original measurement, reproduced from Fig. 3c. The grey trace is a second measurement, carried out under the same conditions and with a fresh sample, to verify repeatability. The red trace is the calculated spectrum (Supplementary Note 8), obtained using the molecular parameters measured in bulk NMR at high field. **b.** High-field (reference proton frequency $f_{\text{ref}} = 500$ MHz) NMR amplitude spectra of ethyl formate.

Spectral constants extracted from these data were used to calculate the low-field spectrum in **a**. The top-left panel shows the full spectrum; the other panels show zoomed-in regions of the spectrum corresponding to one chemical-shift group. The blue circles are the recorded data; the red lines are fits to sums of Lorentzian line shapes used to extract the molecular parameters. Because the expected triplet for the isolated proton (group III) is unresolved, we used the largest J -coupling consistent with the data. The dotted grey lines show the underlying triplet line shapes.



Extended Data Figure 8 | Comparison of NV-ensemble CASR to micrometre-scale inductive NMR detector technologies. Figure adapted from ref. 33 with permission of The Royal Society of Chemistry (<https://doi.org/10.1039/C2SM26065D>). Two additional inductive NMR data points from more recent studies have also been included^{25,34}. The limit of detection is defined as the minimum number of nuclear spins in the sample volume needed to obtain a signal-to-noise ratio of 3 in 1 s of averaging. The sensitivity of the inductive detectors is scaled to a common bias field of $B_0 = 14.1$ T, according to the convention of ref. 33. The CASR

NMR sensitivity for our experiments (large red square) is calculated from the glycerol FNP measurements at $B_0 = 0.088$ T, without scaling the bias field. Projected CASR sensitivities (small red squares) are calculated for both the 10- μl measurement volume at $B_0 = 3$ T and a scaled-up sensor with a measurement volume of about 10 nL (Supplementary Note 9), also at $B_0 = 3$ T. Realizing NV-ensemble CASR measurements at even higher bias fields would be very challenging technically owing to the large NV Rabi frequencies required; we therefore do not extrapolate CASR NMR to $B_0 = 14.1$ T for this comparison.

Redox-influenced seismic properties of upper-mantle olivine

C. J. Cline^{1†}, U. H. Faul², E. C. David^{1†}, A. J. Berry¹ & I. Jackson¹

Lateral variations of seismic wave speeds and attenuation (dissipation of strain energy) in the Earth's upper mantle have the potential to map key characteristics such as temperature, major-element composition, melt fraction and water content^{1–3}. The inversion of these data into meaningful representations of physical properties requires a robust understanding of the micromechanical processes that affect the propagation of seismic waves^{2,3}. Structurally bound water (hydroxyl) is believed to affect seismic properties^{2,3} but this has yet to be experimentally quantified. Here we present a comprehensive low-frequency forced-oscillation assessment of the seismic properties of olivine as a function of water content within the under-saturated regime that is relevant to the Earth's interior. Our results demonstrate that wave speeds and attenuation are in fact strikingly insensitive to water content. Rather, the redox conditions imposed by the choice of metal sleeving, and the associated defect chemistry, appear to have a substantial influence on the seismic properties. These findings suggest that elevated water contents are not responsible for low-velocity or high-attenuation structures in the upper mantle. Instead, the high attenuation observed in hydrous and oxidized regions of the upper mantle (such as above subduction zones) may reflect the prevailing oxygen fugacity. In addition, these data provide no support for the hypothesis whereby a sharp lithosphere–asthenosphere boundary is explained by enhanced grain boundary sliding in the presence of water.

Lattice defects and the structure and composition of grain boundaries in olivine strongly influence the stiffness and strength of the mantle from infinitesimal strains (elastic/anelastic)^{4–6}, to large-strain viscous rheology⁷. Central to experimental efforts in quantifying these effects is understanding the influence of hydrogen-related defects on both diffusional and dislocation-governed processes^{8,9}. It has been suggested, mainly by analogy with observations of water-enhanced viscous deformation of olivine, that the potential seismological importance of water in the mantle might involve either an increase in defect mobility through increased vacancy populations, or water-enhanced diffusivities facilitating grain boundary sliding^{2,10,11}. Substantially modified seismic properties in the presence of water have been observed previously at low frequencies, but only in a single exploratory study conducted under water-saturated conditions¹². Accordingly, hypotheses concerning water-enhanced grain boundary sliding as an explanation for the sharp lithosphere–asthenosphere boundary^{10,11}, and the more general seismological mapping of mantle water contents², remain to be experimentally tested.

Water is conventionally introduced into experimental specimens through the use of solid buffers containing hydrated phases⁹, or as excess fluid, ensuring high water fugacity ($f_{\text{H}_2\text{O}}$) and water-saturated conditions during mechanical testing^{9,12}. However, free fluid phases are not expected under upper-mantle conditions. Conducting experiments under a more relevant water-undersaturated environment requires conditions sufficiently oxidizing to ensure that $f_{\text{H}_2\text{O}} > f_{\text{H}_2}$ (see Methods), for the decoration of intrinsic and extrinsic defects with H^+ .

A recent experimental campaign¹³ used the energetically favoured Ti-clinohumite-like defect (Ti-OH^+)^{14–16} (Ti located on a metal site, charge balanced by a doubly protonated Si vacancy) to assess the sensitivity of large-strain creep in olivine to small (and undersaturated) water concentrations. A conspicuous modification of large-strain rheology was observed in the presence of low concentrations of this defect, with a near-linear enhancement of strain rate in dislocation creep occurring as a function of increasing hydroxyl content, inferred to be due to the increased concentration of associated Si vacancies¹³.

Motivated by the scant experimental data on the influence of water on seismic properties, here we build upon the work of ref. 13 to explore the micro-strain anelastic effects of changes in f_{O_2} and $f_{\text{H}_2\text{O}}$ that are associated with water-undersaturated conditions. We isostatically hot-pressed, and then mechanically tested under a range of f_{O_2} conditions, eight polycrystalline olivine specimens containing different types of hydrated defects in various concentrations. We used different metal sleeves surrounding the olivine specimens during testing to vary the imposed redox conditions¹³—a Pt sleeve to create relatively oxidizing conditions, a $\text{Ni}_{70}/\text{Fe}_{30}$ ('NiFe') sleeve to produce more-reducing conditions, or a Ni sleeve for an intermediate f_{O_2} . The sample suite consisted of undoped and Ti-doped solgel olivine of Fo_{90} composition (containing 90% forsterite)¹³, Ti-doped forsterite and a reconstituted San Carlos olivine (see Extended Data Table 1 for sample compositions). Each specimen was mechanically tested in torsional forced oscillation at periods representative of the seismic band (1–1,000 s), using a confining pressure of 200 MPa and temperatures up to 1,200 °C (see Methods for details).

Fourier-transform infrared (FTIR) spectroscopy was used to identify the hydrous defect and quantify the water content of each specimen after mechanical testing. The FTIR spectra (Fig. 1a) of Ti-bearing specimens are dominated by absorption bands at $3,572\text{ cm}^{-1}$ and $3,525\text{ cm}^{-1}$, attributed to the Ti-OH defect^{14,15}, for which a site-specific calibration factor¹⁶ was used to infer the concentration of chemically bound hydroxyl. The residual broadband absorbance is attributed to the presence of molecular water, and higher concentrations of molecular water are prominent in Ti-free specimens. Additional hydrated defects associated with Si vacancies (at $3,612\text{ cm}^{-1}$) and trivalent ions (at $3,350\text{ cm}^{-1}$) can also be observed in some specimens. The specimen tested within a NiFe-sleeve shows no absorbance within the range $4,000\text{--}3,000\text{ cm}^{-1}$, indicating the absence of both chemically bound hydroxyl and molecular water.

Mechanical data from all specimens exhibit anelastic relaxation typical of the 'absorption band': a monotonic increase of dissipation Q^{-1} and decrease of shear modulus G with increasing oscillation period (Fig. 1b, c) and with increasing temperature. Further, the (G , Q^{-1}) data for each specimen can be adequately described by a Burgers-type creep function¹⁷. Previous forced oscillation results for a Pt-encapsulated and water-saturated dunite¹² show a broad similarity in shear modulus and dissipation to the similarly sleeved, but water-undersaturated, specimens of this study. Additionally, the global Burgers model¹⁷ for a suite

¹Research School of Earth Sciences, Australian National University, Canberra, Australian Capital Territory, Australia. ²Earth Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [†]Present address: Department of Earth Sciences, University College London, London, UK.

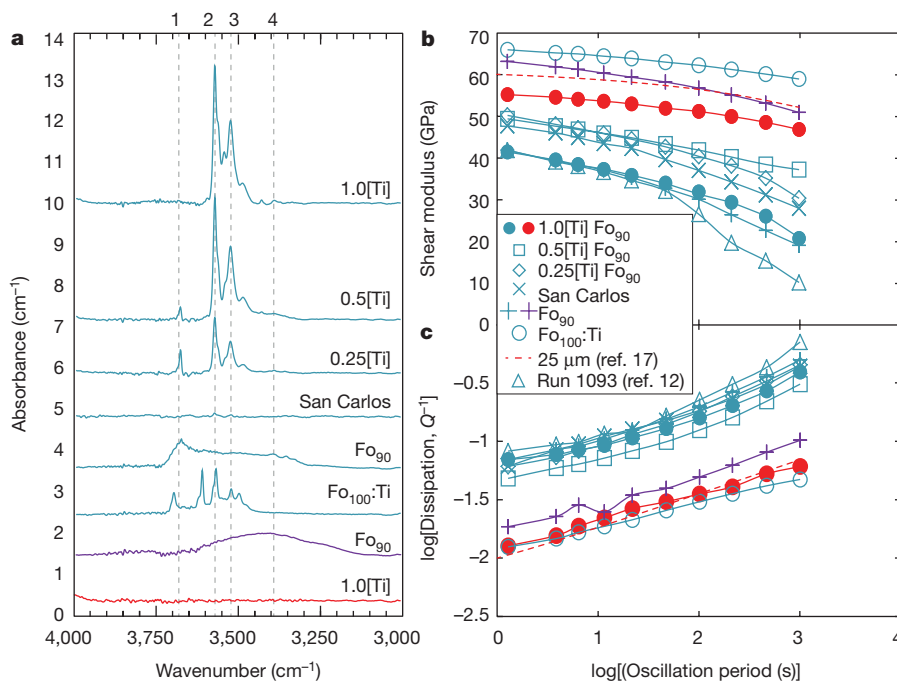


Figure 1 | Seismic properties of all specimens. FTIR spectra for all specimens are shown in **a**, with the corresponding shear modulus in **b** and dissipation data in **c** for a representative temperature of 1,100 °C. Pt-sleeved specimens are represented in blue, Ni-sleeved in purple and NiFe-sleeved in red. In **a**, peak (1) represents a secondary hydrous phase at 3,690 cm⁻¹, peaks (2) and (3) show Ti-OH¹⁴⁻¹⁶ and peak (4) shows OH associated with trivalent ions^{14,16}. In **b** and **c**, the Burgers model for dry olivine¹⁷ was evaluated for a grain size of 25 μm, and is shown by

of dry and melt-free Fo₉₀ polycrystals, updated to fit additional more recent data, closely matches those for the Ti-doped NiFe-sleeved sample tested here, indicating that the presence of the Ti-dopant alone does not affect the observed anelastic relaxation (Fig. 1b).

The dissipation and the associated frequency dependence (dispersion) of the shear modulus are much greater for the Fe-bearing olivine specimens tested within Pt sleeves than for the specimen tested within NiFe (Fig. 1b, c). However, within the suite of Pt-sleeved Fe-bearing polycrystals, the mechanical behaviour is strikingly consistent for specimens ranging widely in concentrations of both bound hydroxyl associated with Ti and molecular water (Fig. 2a). We therefore conclude that the anelastic behaviour of olivine does not vary systematically with either [H/Si], or with Si vacancies associated with the Ti-OH defect, in contrast to the findings of ref. 13 regarding water-enhanced large-strain dislocation creep. In addition, the presence of varying quantities of molecular water (Fig. 2b) and other hydrated defect species also appears to have no effect on the measured seismic properties. Finally, despite the presence of structural water associated with Ti/Mg substitution and Si vacancies, the mechanical properties of the Ti-doped forsterite sample are indistinguishable from those of dry, NiFe-sleeved Fo₉₀ olivine (Figs 1 and 2)—suggesting that in the absence of Fe, higher f_{O_2} and f_{H_2O} conditions are ineffective in enhancing anelastic relaxation.

The magnitude of anelastic relaxation of Fe-bearing olivine presented in Fig. 1b and c instead correlates well with the prevailing f_{O_2} , influenced by the respective metal sleeving materials. The f_{O_2} conditions were inferred from separate hot-pressing experiments, involving the measurement of Fe partitioning between olivine and widely dispersed fine-grained Pt blebs, and vary between Pt- and NiFe-sleeved solgel olivine specimens by 1.7 logarithmic units. The calculated values of f_{O_2} within our large specimens differ from those of the respective metal-oxide buffers, but the expected relative ordering is maintained for the different sleeving materials: for example, Pt > Ni > NiFe (see Methods). Additionally, the f_{O_2} of Pt-sleeved San Carlos olivine is 0.9 logarithmic

units higher than pure solgel olivine sleeved within Pt owing to the presence of impurities, such as Ni and Cr¹⁸. Using the commonly invoked defect charge balance between the concentrations of ferric iron, [Fe³⁺], and metal-site vacancies, [V_M]^{19,20}, we find that Pt-sleeved specimens contain concentrations of such defects roughly twice as high as in NiFe-sleeved specimens. The dissipation measured for the Fe-bearing olivine specimens across the entire range of temperatures and oscillation periods is consistent with the relation $Q^{-1} \approx (f_{O_2})^{1/3}$ (Fig. 3). Our data clearly demonstrate a previously unrecognized link between Q^{-1} and redox conditions. The strength of the inferred

units higher than pure solgel olivine sleeved within Pt owing to the presence of impurities, such as Ni and Cr¹⁸. Using the commonly invoked defect charge balance between the concentrations of ferric iron, [Fe³⁺], and metal-site vacancies, [V_M]^{19,20}, we find that Pt-sleeved specimens contain concentrations of such defects roughly twice as high as in NiFe-sleeved specimens. The dissipation measured for the Fe-bearing olivine specimens across the entire range of temperatures and oscillation periods is consistent with the relation $Q^{-1} \approx (f_{O_2})^{1/3}$ (Fig. 3). Our data clearly demonstrate a previously unrecognized link between Q^{-1} and redox conditions. The strength of the inferred

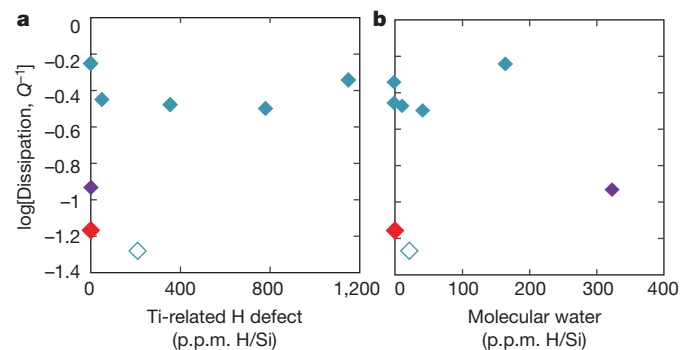


Figure 2 | Dissipation Q^{-1} measured at representative conditions of 1,100 °C and 1,000 s oscillation period. **a**, Q^{-1} versus the corresponding concentration in atom parts per million (p.p.m.) of Ti-related H/Si. **b**, Q^{-1} versus the corresponding concentration of molecular water. Colours are as in Fig. 1, with Fe-bearing olivine represented by solid diamonds, and the Ti-doped forsterite specimen represented by the hollow diamond. Dissipation data are normalized to a common grain size of 25 μm. Errors on measurements of dissipation, given by $\sigma[\log(Q^{-1})] = 0.05$ and water content (standard deviation) are smaller than the symbols used for plotting.

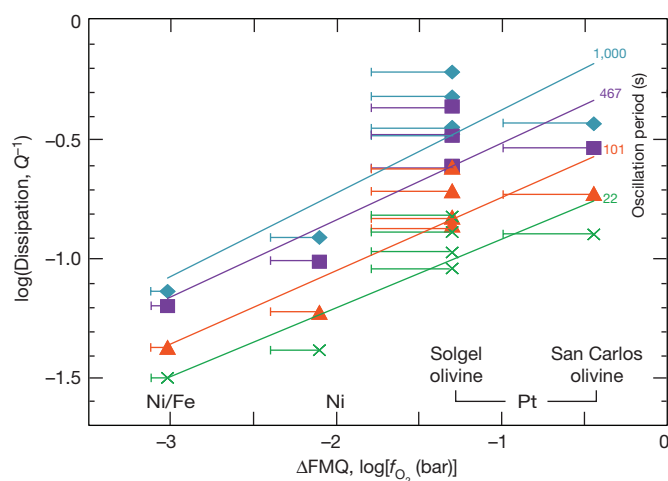


Figure 3 | Dissipation data at 1,100 °C plotted as a function of f_{O_2} for different capsule materials and several representative oscillation periods. Linear fits of the data at the various oscillation periods (coloured numbers) have an average slope of 0.31. Dissipation data are normalized to a common grain size of 25 μm . The lines attached to plotting symbols provide an indication of the combined uncertainty³⁰ arising from analytical errors and the use of alternative thermodynamic data (see Methods). f_{O_2} is given as the deviation from that of the fayalite–magnetite–quartz (FMQ) equilibrium. Distinct f_{O_2} values for Pt-sleeved solgel-derived and San-Carlos-derived olivine specimens are shown.

sensitivity to the variation of f_{O_2} results from large changes in Q^{-1} for the relatively small range of f_{O_2} .

Increased concentrations and mobilities of lattice defects resulting from more oxidizing conditions (specifically Fe^{3+} and vacancies on metal sites) and associated increases in lattice and grain-boundary diffusivities are probably the cause of the observed enhancement in anelastic relaxation. Accordingly, under these more oxidizing conditions, the effective grain-boundary viscosity will be reduced, and hence so will the characteristic timescales τ_e and τ_M (the Maxwell time), which are associated with elastically and diffusively assisted/accommodated grain-boundary sliding, respectively⁵. Indeed, values of τ_M determined from the Burgers model fitting¹⁷ of (G, Q^{-1}) data for each specimen vary systematically with f_{O_2} (Fig. 4). It has been suggested that water contents of 0.1 wt% may decrease τ_e by two to four orders of magnitude and thus enhance dissipation for elastically accommodated grain-boundary sliding, potentially explaining the sharp lithosphere–asthenosphere boundary^{10,11}. Here we have demonstrated that there exists no such sensitivity of seismic properties to the presence of water under our

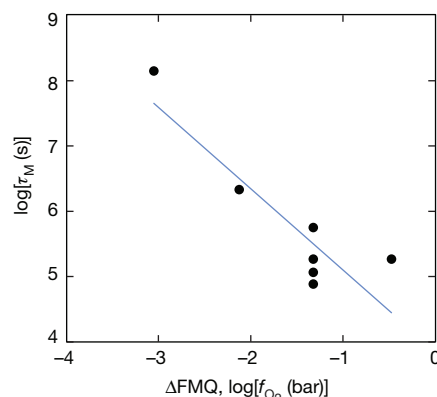


Figure 4 | Values of $\log(\tau_M)$ from the refined Burgers model of each Fo_{90} olivine specimen plotted as a function of f_{O_2} . The data are best-fitted with a line of slope -1.2 .

experimental conditions, but we do observe a comparable enhancement of anelastic relaxation attributable to an increase of f_{O_2} by 1.7 logarithmic units and the concomitant defects.

Results from the recent NoMelt seismic experiment, which was conducted in 60–80-million-year-old lithosphere in the central Pacific, away from ridges and hotspots²¹, are consistent with our findings of a negligible influence of water on seismic properties. Analyses of basalts from the East Pacific Rise and Macquarie Island indicate that water contents for the Pacific upper mantle approach 200 parts per million by weight H_2O (ref. 22, and references therein). Taking partitioning of the bulk water content between olivine and pyroxenes into account, olivine in the upper mantle may thus contain approximately 400–1,000 atom parts per million H/Si ^{23,24} (that is, within the range experimentally investigated here). In addition, electrical conductivity measurements collected as part of the NoMelt experiment indicate increased conductivity in the asthenosphere, consistent with water contents of 25–400 parts per million by weight²⁵. The presence of this amount of water was previously expected to reduce seismic velocities substantially^{2,10}. However, the results of the NoMelt survey²¹ indicate that the observed velocity structure, including the velocity minimum, can be explained with a relatively minor additional enhancement of the anelastic relaxation characteristics of dry and relatively reduced olivine¹⁷, down to a depth of 300 km, that is, well below the melt (and volatile) extraction horizon (Fig. 5).

In contrast, the low velocity and high attenuation observed above subducting slabs are typically attributed to volatile release (and therefore higher water contents²⁶) and to the presence of melt³. However,

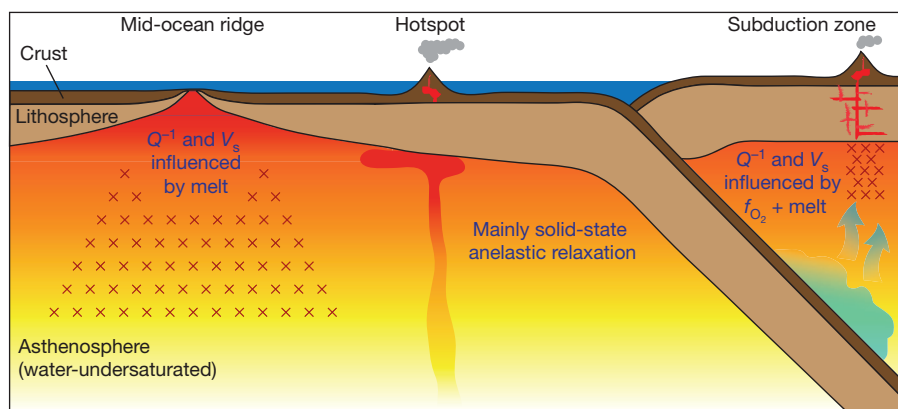


Figure 5 | Dominant influences on anelastic relaxation responsible for the reduced seismic shear-wave velocities and attenuation of seismic waves. ‘Normal’ asthenosphere beneath old oceanic lithosphere, which is water-undersaturated and contains only a very minor melt fraction, is subject mainly to solid-state anelastic relaxation. Higher melt fractions in

mid-ocean ridge environments result in enhanced relaxation through the partial wetting of grain boundaries (red crosses indicate the presence of appreciable melt), whereas mantle-wedge environments are influenced by both the presence of melt and increased oxygen fugacity, f_{O_2} . Blue text lists the influences on seismic shear-wave velocities V_s and attenuation.

measurements of the ratio between ferric iron and total iron ($\text{Fe}^{3+}/\Sigma\text{Fe}$) in arc lavas indicate that f_{O_2} is also higher by 1–1.5 logarithmic units in the mantle wedge in comparison to adjacent oceanic lithosphere^{27–29}. With the $Q^{-1} \approx (f_{\text{O}_2})^{1/3}$ result of this study, such variations of f_{O_2} are predicted to enhance dissipation 2–3-fold, with the associated dispersion reflected in reduced wave speeds. Therefore, although water content generally correlates with magmatic ratios²⁷ of $\text{Fe}^{3+}/\Sigma\text{Fe}$, and may also increase grain size³, the reduced velocities and increased attenuation above subducting slabs may be attributable not to the presence of fluids or hydrated defects, but rather to the prevailing redox conditions and the presence of melt (Fig. 5).

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 August 2017; accepted 4 January 2018.

- Romanowicz, B. A. & Mitchell, B. J. in *Treatise on Geophysics* 2nd edn, 789–827 (Elsevier, 2015).
- Karato, S. Mapping water content in upper mantle. In *Inside the Subduction Factory* Vol. 138, 135–152 (Geophysical Monograph Series, AGU, 2003).
- Abers, G. et al. Reconciling mantle attenuation-temperature relationships from seismology, petrology, and laboratory measurements. *Geochem. Geophys.* **15**, 3521–3542 (2014).
- Jacobsen, S. D. et al. Effects of hydration on the elastic properties of olivine. *Geophys. Res. Lett.* **35**, L14303 (2008).
- Jackson, I. in *Treatise on Geophysics* 2nd edn, 539–571 (Elsevier, 2015).
- Farla, R. J., Jackson, I., Fitz Gerald, J. D., Faul, U. H. & Zimmerman, M. E. Dislocation damping and anisotropic seismic wave attenuation in Earth's upper mantle. *Science* **336**, 332–335 (2012).
- Bai, Q., Mackwell, S. J. & Kohlstedt, D. L. High-temperature creep of olivine single crystals 1. Mechanical results for buffered samples. *J. Geophys. Res.* **96**, 2441 (1991).
- Karato, S.-I. & Jung, H. Effects of pressure on high-temperature dislocation creep in olivine. *Phil. Mag.* **83**, 401–414 (2003).
- Mei, S. & Kohlstedt, D. Influence of water on plastic deformation of olivine aggregates. 2. Diffusion creep regime. *J. Geophys. Res.* **105**, 21457–21469 (2000).
- Karato, S.-I. On the origin of the asthenosphere. *Earth Planet. Sci. Lett.* **321/322**, 95–103 (2012).
- Olugboji, T. M., Karato, S.-I. & Park, J. Structures of the oceanic lithosphere-asthenosphere boundary: mineral-physics modeling and seismological signatures. *Geochem. Geophys.* **14**, 880–901 (2012).
- Aizawa, Y. et al. Seismic properties of Anita Bay dunite: an exploratory study of the influence of water. *J. Petrol.* **49**, 841–855 (2008).
- Faul, U. H., Cline, C. J. II, David, E. C., Berry, A. J. & Jackson, I. Titanium-hydroxyl defect-controlled rheology of the Earth's upper mantle. *Earth Planet. Sci. Lett.* **452**, 227–237 (2016).
- Berry, A. J., Hermann, J., O'Neill, H. S. C. & Foran, G. J. Fingerprinting the water site in mantle olivine. *Geology* **33**, 869 (2005).
- Walker, A. M., Hermann, J., Berry, A. J. & O'Neill, H. S. C. Three water sites in upper mantle olivine and the role of titanium in the water weakening mechanism. *J. Geophys. Res.* **112**, B05211 (2007).
- Kovács, I., O'Neill, H. S. C., Hermann, J. & Hauri, H. Site-specific infrared OH absorption coefficients for water substitution into olivine. *Am. Mineral.* **95**, 292–299 (2010).
- Jackson, I. & Faul, U. H. Grain-size sensitive viscoelastic relaxation in olivine: Towards a robust laboratory-based model for seismological application. *Phys. Earth Planet. Inter.* **183**, 151–163 (2010).
- De Hoog, J., Gail, L. & Cornell, D. Trace-element geochemistry of mantle olivine and application to mantle petrogenesis and geothermobarometry. *Chem. Geol.* **270**, 196–215 (2010).
- Nakamura, A. & Schmalzried, H. On the nonstoichiometry and point defects of olivine. *Phys. Chem. Mineral.* **10**, 27–37 (1983).
- Dohmen, R. & Chakraborty, S. Fe–Mg diffusion in olivine II: point defect chemistry, change of diffusion mechanisms and a model for calculation of diffusion coefficients in natural olivine. *Phys. Chem. Mineral.* **34**, 409–430 (2007).
- Lin, P.-Y. P. et al. High-resolution seismic constraints on flow dynamics in the oceanic asthenosphere. *Nature* **535**, 538–541 (2016).
- Shimizu, K. et al. Two-component mantle melting-mixing model for the generation of mid-ocean ridge basalts: implications for the volatile content of the Pacific upper mantle. *Geochim. Cosmochim. Acta* **176**, 44–80 (2016).
- Hirth, G. & Kohlstedt, D. L. Water in the oceanic upper mantle: implications for rheology, melt extraction and the evolution of the lithosphere. *Earth Planet. Sci. Lett.* **144**, 93–108 (1996).
- Tenner, T. J., Hirschmann, M. M., Withers, A. C. & Ardia, P. H₂O storage capacity of olivine and low-Ca pyroxene from 10 to 13 GPa: consequences for dehydration melting above the transition zone. *Contrib. Mineral. Petrol.* **163**, 297–316 (2012).
- Sarafian, E. et al. The electrical structure of the central Pacific upper mantle constrained by the NoMelt experiment. *Geochem. Geophys.* **16**, 1115–1132 (2015).
- Stachnik, J. C., Abers, G. A. & Christensen, D. H. Seismic attenuation and mantle wedge temperatures in the Alaska subduction zone. *J. Geophys. Res.* **109**, B10304 (2004).
- Kelley, K. A. & Cottrell, E. Water and the oxidation state of subduction zone magmas. *Science* **325**, 605–607 (2009).
- Brounce, M. N., Kelley, K. A. & Cottrell, E. Variations in $\text{Fe}^{3+}/\Sigma\text{Fe}$ of Mariana arc basalts and mantle wedge f_{O_2} . *J. Petrol.* **55**, 2513–2536 (2014).
- McCammon, C. A. et al. Oxidation state of iron in hydrous mantle phases: implications for subduction and mantle oxygen fugacity. *Phys. Earth Planet. Inter.* **143–144**, 157–169 (2004).
- Faul, U. H., Cline, C. J. II, Berry, A., Jackson, I. & Garapic, G. Constraints on oxygen fugacity within metal capsules. *Phys. Chem. Mineral.* <https://doi.org/10.1007/s00269-017-0937-7> (2017).

Acknowledgements We thank H. Kokkonen and H. Miller for technical support with sample preparation and operation of the Australian National University rock physics laboratory, H. Chen for assistance with electron backscatter diffraction, and S. Karato of Yale University for suggestions that improved the manuscript. This work was supported by grant DP130103848 from the Australian Research Council to I.J., A.J.B., U.H.F. and S. Karato. U.H.F. acknowledges support from NSF grant EAR 1321889. C.J.C. gratefully acknowledges funding by an Australian National University International PhD Research Scholarship.

Author Contributions C.J.C. conducted the forced-oscillation measurements (with contributions from E.C.D.). C.J.C. obtained the microstructural and spectroscopic data, subsequently interpreted by U.H.F., E.C.D., I.J., A.J.B. and C.J.C. f_{O_2} and Fe^{3+} calculations were performed by U.H.F. The manuscript was written by C.J.C. with contributing revisions and discussion from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to C.J.C. (christopher.cline@anu.edu.au).

Reviewer Information *Nature* thanks G. Abers and T. Irifune for their contribution to the peer review of this work.

METHODS

Specimen fabrication. Three types of starting materials were used to produce the eight olivine specimens tested in this study. Seven specimens were fabricated from precursor powder prepared by a solution-gelation process (two undoped Fo_{90} , four Ti-doped Fo_{90} and one Ti-doped Fo_{100}), and the other from powder produced by crushing hand-picked San Carlos olivine phenocrysts. The fabrication procedure for the different solgel-derived olivine powders follows the methodology of ref. 31 for Fo_{90} and of ref. 13 for the Ti-doped forsterite material. Nitrates of Fe and Mg (and tetra-ethoxy orthotitanate for the Ti-bearing material) were mixed with a nonstoichiometric amount of tetraethyl orthosilicate to ensure that the final olivine was orthopyroxene-saturated. Gelation of this mix was initiated with a small amount of HNO_3 , followed by dehydration, and firing in air within a Pt crucible to remove all nitrates. The resulting powders were subjected to multiple firings in a 1-atm furnace in a 50:50 $\text{CO}-\text{CO}_2$ gas mix at 1,400 °C for 16 h, with intervening grindings to produce a pure and homogeneous starting material with an average grain diameter of about 1 μm . The San Carlos precursor material is identical to that fabricated and used by ref. 32.

The precursor powder was then cold-pressed at about 10 MPa into a set of five 3-g pellets that were fired in an atmosphere of 50:50 $\text{CO}-\text{CO}_2$ gas for 16 h at either 1,400 °C for the solgel-derived or 1,300 °C for the San-Carlos-derived material. The stack of five pellets was accommodated within a laser-welded Pt capsule, sandwiched between alumina pistons within a thin-walled mild steel jacket and hot-pressed at 1,200 °C and 300 MPa for 24 h in a Paterson internally heated gas-medium pressure vessel. Specimens were recovered by dissolving the steel jacket in HNO_3 and subsequently peeling off the remaining Pt capsule. Transverse sections were cut about 1.5 mm from the end of each sample for microstructural analysis, with the bulk of the specimens then being precision ground to a right cylindrical shape with a diameter of 11.5 mm and lengths ranging between 31 mm and 33 mm. Each such specimen was then fired at 600 °C in Ar to remove all traces of organic lubrication used in the grinding, and held in a drying oven at 110 °C until mechanical testing. **Mechanical testing.** Mechanical testing of each specimen, via forced torsional oscillation, was conducted using the modified Paterson gas-medium apparatus originally described by ref. 33, and later with updated procedures reported by ref. 34. Specimens were sleeved in either Pt, $\text{Ni}_{70}\text{Fe}_{30}$ or Ni foil and placed inside a mild-steel jacket along with Lucalox alumina torsion rods. Pressure was applied via Ar, and maintained at 200 MPa for the duration of the experiment to load all interfaces (such as those between the sample and the aluminium oxide torsion rods) within the experimental assembly, and thereby to ensure appropriate frictional coupling. The temperature was increased to 1,200 °C, and maintained for a time interval ranging between 23 h and 45 h, during which the mechanical response was monitored until steady and repeatable values of normalized compliance and phase lag were observed, indicating that the evolution of the microstructure and chemical environment is complete. Since this evolution is thermally activated, once steady-state mechanical response is achieved at the highest temperature, continued mechanical testing during the slow staged cooling (over approximately 1 week) is thought to be representative of a single chemical environment and microstructure.

Torsional testing involves the application of a sinusoidal stress at each of ten oscillation periods, logarithmically equi-spaced between 1 s and 1,000 s. Maximum strains at the highest temperature and longest oscillation period were kept to $<10^{-5}$ to ensure linear behaviour, which was confirmed before beginning the formal test at the highest temperature by halving the strain amplitude and monitoring values of Q^{-1} . A complete experiment encompasses testing at oscillation periods of 1–1,000 s followed by a complementary torsional microcreep test, which is intended to assess the recoverability of the viscoelastic strain³³. After both types of torsional tests were complete, the temperature was reduced by 50 °C, and the procedure was repeated until nearly elastic behaviour was observed, upon which the temperature interval was increased to 100 °C, until room temperature was reached.

Water content and grain-size analysis. Water content was determined using FTIR, for both the hot-pressed precursor and for specimens recovered after mechanical testing. Spectra were recorded using unpolarized light from 400- μm -thick traverse sections with an apertured spatial resolution of 200 $\mu\text{m} \times 200 \mu\text{m}$ using a Bruker Hyperion 2000 microscope and a Tensor 27 spectrometer with a MCT-A detector. Samples were placed in a Plexiglas box and purged with dry air for 45 min before, and continually during, spectral acquisition. Each spectrum is the average of 64 scans, recorded between 600 cm^{-1} and 5,000 cm^{-1} with a resolution of 4 cm^{-1} .

Determination of water content follows the method of ref. 13, in which a site-specific calibration factor of $k = 0.18$ is used to convert the integrated area between 3,450 cm^{-1} and 3,600 cm^{-1} into the concentration of crystallographically bound hydroxyl. The residual broadband absorption is attributed to the presence of molecular water¹⁶, and is quantified separately using the magnitude of absorbance at 3,450 cm^{-1} and a molar absorption coefficient of 115 $\text{mol}^{-1} \text{cm}^{-1}$. An unidentified secondary hydrous phase identified in the FTIR spectra of some specimens will not affect the quoted water contents or the measured mechanical properties. Presumably this phase formed during the slow staged cooling, and

because hydrated phases are unstable above 700 °C³⁵ and anelasticity is not appreciable or measurable below this temperature, this hydrous phase will not affect the mechanical response of the specimens. The presence of such a hydrous phase is suggestive of a hydrous environment during mechanical testing, even in the absence of any other structural H-related defects.

Electron back-scatter diffraction lattice orientation maps were collected to determine the grain size of each specimen after mechanical testing. Sections were successively polished down from 800-grit silicon carbide to a final vibratory polish using 0.05 μm colloidal silica. Orientation maps were collected at the Center for Advanced Microscopy at Australian National University using a Zeiss Ultra Plus field-emission scanning electron microscope fitted with an Oxford Nordlys S electron back-scatter diffraction camera. A representative orientation map is available in Extended Data Fig. 1.

Determination of f_{O_2} and Fe^{3+} in olivine. The oxygen fugacity f_{O_2} was determined for the differently sleeved olivine specimens through separate hot-pressing experiments. 1 wt% Pt-black was mixed with either solgel-derived or San-Carlos-derived olivine powder; then the material was encapsulated in either Pt, Ni, NiFe or Fe foil and hot-pressed in a Paterson internally heated gas apparatus at 1,200 °C and 300 MPa for 24 h. The recovered specimens were then analysed using standardized energy-dispersive X-ray spectroscopy and electron microprobe wavelength-dispersive spectroscopy analysis to yield the Fe contents of both the small Pt particles (about 10 μm) and of the surrounding olivine. Using the metal activities γ in both the olivine and alloyed Pt blebs, this then results in a calculable³⁶ f_{O_2} value, using the relation³⁷:

$$\log(f_{\text{O}_2}) = 2\log(\gamma_{\text{Fe}}^{\text{ol}}) - 2\log(\gamma_{\text{Fe}}^{\text{alloy}}) + 2\log(X_{\text{Fe}}^{\text{ol}}/X_{\text{Fe}}^{\text{alloy}}) - \log(a_{\text{SiO}_2}) - \log(K_2)$$

where a_{SiO_2} is silica activity, X_{Fe} is the mole fraction of Fe either in the olivine or in the Pt blebs and K_2 is the appropriate equilibrium constant. The activity of Fe in olivine is calculated as in ref. 37:

$$\ln(\gamma_{\text{Fe}}^{\text{ol}}) = (1 - X_{\text{Fe}}^{\text{ol}})^2(600 + 0.0013P)/T$$

where P is pressure and T is absolute temperature. The activity of Fe in the PtFe metal alloy is calculated from ref. 36:

$$\ln\gamma_{\text{Fe}}^{\text{alloy}} = [W_{\text{G1}} + 2(W_{\text{G2}} - W_{\text{G1}})X_{\text{Fe}}^{\text{alloy}}](X_{\text{Fe}}^{\text{alloy}})^2/RT$$

where $W_{\text{G1}} = 138 \text{ kJ mol}^{-1}$ and $W_{\text{G2}} = 90.8 \text{ kJ mol}^{-1}$ are the Margules parameters and R is the gas constant. An earlier model for the activity of Fe in the alloy^{37,38} resulted in lower calculated values of f_{O_2} , particularly for oxidizing conditions; these uncertainties are indicated by the lines in Fig. 3. A detailed analysis of these additional hot-pressing experiments designed for the determination of f_{O_2} in large solid samples is available in ref. 30.

Using equations (3) to (5) of ref. 20 (see also references therein), in combination with the appropriate f_{O_2} as calculated above, we can determine the concentrations of Fe^{3+} (that is, $[\text{Fe}_M^*]$) in our Pt- and NiFe-sleeved specimens. Using the charge neutrality conditions of olivine, we can also determine the concentration of the metal-site vacancies $[\text{V}_M'']$ in our Pt- and NiFe-sleeved specimens. Olivine at Si-saturated conditions (coexisting with pyroxene) possesses a concentration of Fe^{3+} defects described by:

$$\log[\text{Fe}_M^*] = \frac{1}{6}(\log K_x + 2\log 2 + 4\log X_{\text{Fa}}^{\text{ol}} + \log f_{\text{O}_2} + \log a_{\text{SiO}_2})$$

where $\log K_x = -7.32 - (90,000/2.303RT)$ (ref. 19), T is in kelvin and X_{Fa} is fayalite content. From this calculation of $[\text{Fe}_M^*]$, the ratio of Fe^{3+} to total Fe can be accessed via the relation:

$$\text{Fe}^{3+}/\Sigma\text{Fe} = \frac{[\text{Fe}_M^*]}{2X_{\text{Fa}}^{\text{ol}}}$$

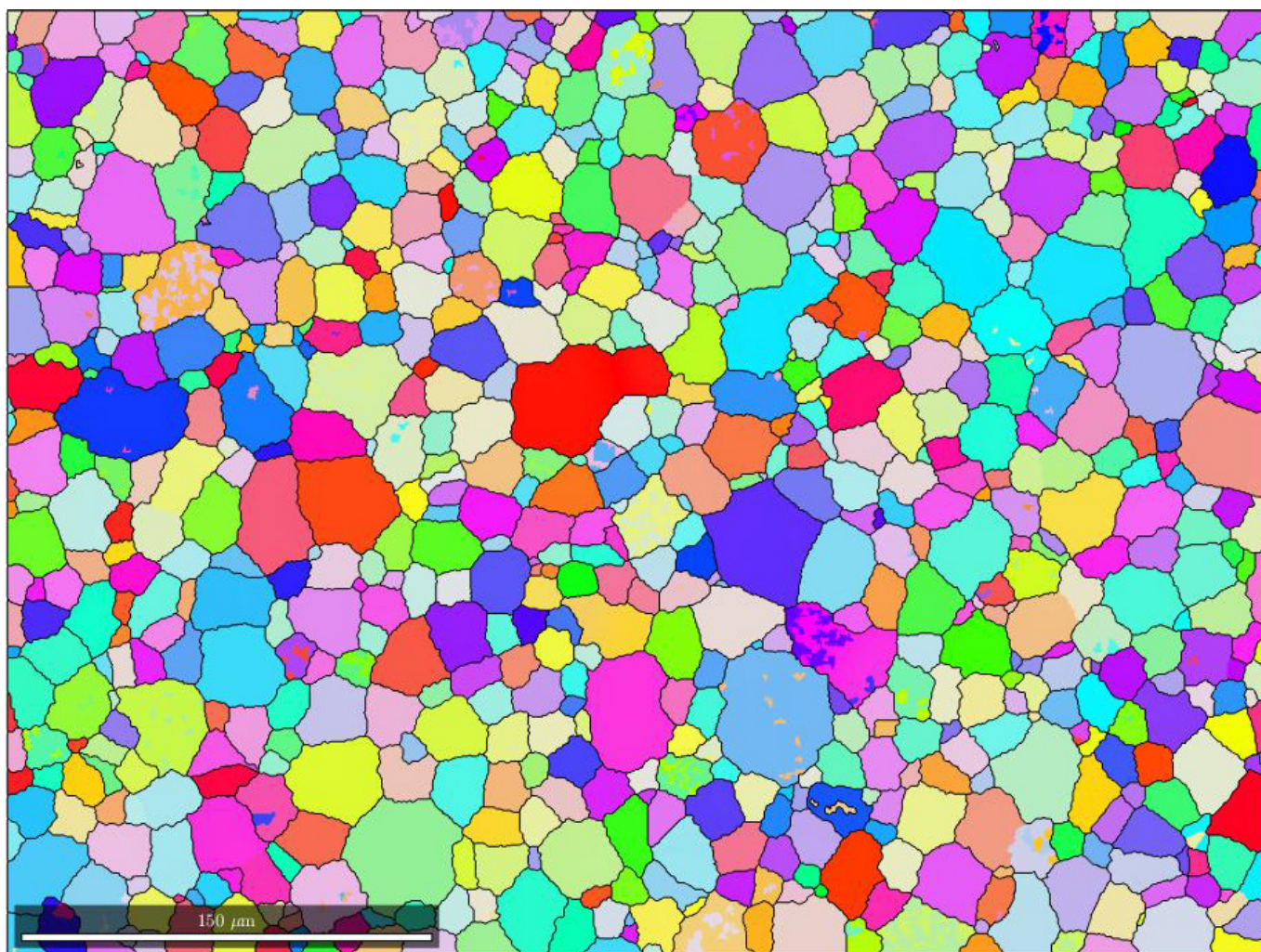
The calculated Fe^{3+} percentage of total Fe for Pt-sleeved olivine is 0.17 in the interior. NiFe-sleeved specimens exhibit roughly half of this latter value at $\text{Fe}^{3+} = 0.09\%$, and Ni-sleeved olivine is intermediate with 0.13%.

Hydrogen speciation and water retention as a function of redox conditions. Water present in our forced-oscillation experiments is not intentionally added. The source of water is presumed to be either via adsorption of moisture within the specimen after removal from the gas mixing furnace and before pressurization or via ingress of H from the relatively reducing furnace environment into the more oxidizing environment within the specimen. Given the evident availability of H in our experiments, the presence of hydrated defects can be controlled through the use of different metal-sleeving materials, imposing either favourable or unfavourable redox conditions for the oxidation of H.

The relative fugacities of H_2 and H_2O are controlled by the equilibrium $H_2 + 1/2O_2 = H_2O$ with an equilibrium constant K_w of 5.9 at 1,200 °C (ref. 39). Using the values of f_{O_2} determined from the separate hot-pressing experiments detailed above, the difference of +1.7 logarithmic units in f_{O_2} between NiFe- and Pt-sleeved olivine thus increases the value of $f_{H_2O}/f_{H_2} = K_w(f_{O_2})^{1/2}$ from only 2.4 to 13.5 (where f_{O_2} is in units of bar), which is consistent with the inferred dominance of oxidized hydrogen within Pt sleeves, and the loss of water as hydrogen from NiFe-sleeved olivine.

Data availability. Source data for Figs 1–4 are provided with the online version of the paper.

31. Jackson, I. *et al.* Grain-size-sensitive seismic wave attenuation in polycrystalline olivine. *J. Geophys. Res.* **107**, 2360 (2002).
32. Tan, B. H., Jackson, I. & Fitz Gerald, J. D. High-temperature viscoelasticity of fine-grained polycrystalline olivine. *Phys. Chem. Miner.* **28**, 641–664 (2001).
33. Jackson, I. & Paterson, M. S. A high pressure, high temperature apparatus for studies of seismic wave dispersion and attenuation. *Pure Appl. Geophys.* **141**, 445–466 (1993).
34. Jackson, I., Barnhoorn, A., Aizawa, Y. & Saint, C. Improved procedures for the laboratory study of high-temperature viscoelastic relaxation. *Phys. Earth Planet. Inter.* **172**, 104–115 (2009).
35. Engi, M. & Lindsley, D. H. Stability of titanian clinohumite: Experiments and thermodynamic analysis. *Contrib. Mineral. Petrol.* **72**, 415–424 (1980).
36. Kessel, R., Beckett, J. R. & Stolper, E. M. Thermodynamic properties of the Pt-Fe system. *Am. Mineral.* **86**, 1003–1014 (2001).
37. Rubie, D. C., Karato, S., Yan, H. & O'Neill, H. S. C. Low differential stress and controlled chemical environment in multianvil high-pressure experiments. *Phys. Chem. Mineral.* **20**, 315–322 (1993).
38. Heald, E. Thermodynamics of iron-platinum alloys. *Trans. Metal. Soc. AIME* **239**, 1337–1340 (1967).
39. Paterson, M. The thermodynamics of water in quartz. *Phys. Chem. Mineral.* **13**, 245–255 (1986).



Extended Data Figure 1 | Electron back-scatter diffraction map of sample 1623 (0.5[Ti]) after mechanical testing, coloured to indicate different crystallographic orientations. A near-‘foam’ microstructure

is present, with apparent grain-boundary serrations being an artefact of the post-processing grain boundary reconstruction using MTEX software (<http://mtex-toolbox.github.io/>).

Extended Data Table 1 | Summary of sample characteristics

Sample	Precursor material	Normalized [Ti] ^a	Ti content	Metal sleeve	Bound hydroxyl ^c	Grain size, μm^{d}	$\Delta\text{FMQ}^{\text{e}}$ interior ^e
1515	Fo ₉₀ Solgel	1	802(4) ^b	Pt	1150(190)	25.4(9)	-1.3(5)
1579	Fo ₉₀ Solgel	1	802(4) ^b	Ni ₇₀ Fe ₃₀	0	25.0(10)	-3.0(1)
1623	Fo ₉₀ Solgel	0.5	396(1) ^b	Pt	780(36)	18.6(6)	-1.3(5)
1637	Fo ₉₀ Solgel	0.25	176(1) ^b	Pt	355(32)	19.9(8)	-1.3(5)
1646	San Carlos	N/A	18-46 ^f	Pt	50(17)	80.9(20)	-0.4(7)
1651	Fo ₉₀ Solgel	0	0	Pt	0	20.6(5)	-1.3(5)
1684	Fo ₁₀₀ Solgel	1	800(5)	Pt	200(7)	8.2(6)	-0.9
1689	Fo ₉₀ Solgel	0	0	Ni	0	10.7(5)	-2.0(3)

^aSamples are denoted by the nomenclature x[Ti] where x is the nominal amount of Ti normalized to the sample with the highest Ti-dopant concentration. The values of x are given in this column and correspond to the intended composition.

^bAtom parts per million Ti/Si determined by LA-ICPMS. Values in parentheses are standard errors of the mean (%).

^cAtom parts per million H/Si by FTIR in the sample interior, associated with Ti-clinohumite-like defects only. Values in parentheses indicate one standard deviation of data collected on longitudinal maps or radial (linear) transects.

^dCalculated using electron back-scatter diffraction orientation maps in combination with MTEX processing software. Errors represent the raster step size of the orientation map.

^eIn units of $\log[f_{\text{O}_2}$ (bar)]. The combined uncertainty is calculated from analytical errors and the use of alternative thermodynamic data for Fe-bearing specimens; see Methods for additional details.

^fFrom ref. 18.

Altruism in a volatile world

Patrick Kennedy¹, Andrew D. Higginson², Andrew N. Radford¹ & Seirian Sumner³

The evolution of altruism—costly self-sacrifice in the service of others—has puzzled biologists¹ since *The Origin of Species*. For half a century, attempts to understand altruism have developed around the concept that altruists may help relatives to have extra offspring in order to spread shared genes². This theory—known as inclusive fitness—is founded on a simple inequality termed Hamilton's rule². However, explanations of altruism have typically not considered the stochasticity of natural environments, which will not necessarily favour genotypes that produce the greatest average reproductive success^{3,4}. Moreover, empirical data across many taxa reveal associations between altruism and environmental stochasticity^{5–8}, a pattern not predicted by standard interpretations of Hamilton's rule. Here we derive Hamilton's rule with explicit stochasticity, leading to new predictions about the evolution of altruism. We show that altruists can increase the long-term success of their genotype by reducing the temporal variability in the number of offspring produced by their relatives. Consequently, costly altruism can evolve even if it has a net negative effect on the average reproductive success of related recipients. The selective pressure on volatility-suppressing altruism is proportional to the coefficient of variation in population fitness, and is therefore diminished by its own success. Our results formalize the hitherto elusive link between bet-hedging and altruism^{4,9–11}, and reveal missing fitness effects in the evolution of animal societies.

The widespread phenomenon of organisms paying costs to help others (altruism) is a long-standing paradox in biology^{1,2}. Recently, variance-averse investment in stochastic environments (bet-hedging) has been suggested as an explanation for a number of major puzzles in the evolution of altruism, including the origins of sociality in birds^{9,11,12}, insects¹³ and rodents¹⁴, the altitudinal distribution of eusocial species⁷, and the evolution of cooperation between eusocial insect colonies¹⁵. The global distribution of animal societies is linked to environmental stochasticity⁴. In birds^{6,12}, mammals¹⁶, bees⁷ and wasps⁸, cooperation is more common in unpredictable or harsh environments. However, the effects of stochasticity have largely been omitted from social evolutionary theory. There are a few notable exceptions: in ref. 17 it is argued that selection will maximize expected inclusive fitness under uncertainty; ref. 18 shows that mutualism between non-relatives could counteract kin selection by dampening stochasticity; and stochastic effects on reproductive value are explored in ref. 19. However, despite speculation^{11,20}, the proposed link between bet-hedging and altruism⁹ has remained elusive⁴. We resolve this link by presenting a stochastic generalization of Hamilton's rule (stochastic Hamilton's rule), which predicts when organisms should pay a cost to influence the variance in the reproductive success of their relatives.

We allow the environmental state π to fluctuate among the possible states Π ; stochasticity is the condition that states are unpredictable. We follow the established method of capturing fitness effects as regression slopes¹. Both the fitnesses w_x of individual organisms and the average fitness \bar{w} in the population may vary among the states Π . We denote the k th central moment of \bar{w} as $\langle\langle^k \bar{w}\rangle\rangle$. The joint distribution of the fitness of individual x (w_x) and \bar{w} across states Π is captured by their mixed moments (covariance, $k=1$; coskewness, $k=2$; cokurtosis, $k=3$ and so on; Supplementary Information A1). Altruists may not only alter

the expected number of offspring (mean, $k=0$), but also may reduce the variation in offspring number (variance, $k=1$) or increase the likelihood of large numbers of offspring (skew, $k=2$). We denote the effect of the actor on the expected number of offspring of the recipient as the benefit b_μ , the effect of the actor on its own expected number of offspring as the cost c_μ , and relatedness as r . Likewise, we denote the effect of the actor on the k th mixed moment defining the reproductive success of the recipient as b_k , and the effect of the actor on the k th mixed moment of its own reproductive success as c_k . The stochastic Hamilton's rule is therefore:

$$r \cdot \overbrace{\left(b_\mu + \sum_{k=1}^{\infty} \frac{(-1)^k}{\mathbb{E}_\pi[\bar{w}]^k} (\langle\langle^k \bar{w}\rangle\rangle b_\mu + b_k) \right)}^B > \overbrace{\left(c_\mu + \sum_{k=1}^{\infty} \frac{(-1)^k}{\mathbb{E}_\pi[\bar{w}]^k} (\langle\langle^k \bar{w}\rangle\rangle c_\mu + c_k) \right)}^C \quad (1)$$

Empirical tests of Hamilton's rule have looked for benefits and costs that constitute effects on the mean reproductive success of recipients and actors, using the form $rb_\mu > c_\mu$ (henceforth, means-based Hamilton's rule)²¹. However, equation (1) reveals that b_μ is a single component of a range of potential benefits of altruism. Conclusions based on mean reproductive success (b_μ and c_μ) overlook effects on the variance of the distribution from which a recipient samples its reproductive success.

Asocial bet-hedging has been analysed extensively³, and is typically described in terms of costs and benefits: the cost is a reduction in mean reproductive success, whereas the benefit is a reduction in the variance of reproductive success³. Following speculation that these benefits and costs could be accrued by different partners^{9,13}—actors pay costs whereas recipients derive benefits (Fig. 1a)—we refer to decoupled benefits and costs as altruistic bet-hedging. We let b_σ and c_σ denote, respectively, the effects on the standard deviation (volatility) of the recipient and actor in reproductive success (weighted by its correlation with population average reproductive success \bar{w} ; for details see Extended Data Table 1). We introduce the stochasticity coefficient v as the coefficient of variation in \bar{w} across environmental conditions ($v = \frac{\sigma_\pi[\bar{w}]}{\mathbb{E}_\pi[\bar{w}]}$; Fig. 1b). For cases in which the actor can affect both the mean and the volatility (but not higher moments) of the reproductive success of the recipient, equation (1) simplifies (Supplementary Information A2) to:

$$r(b_\mu + vb_\sigma) > c_\mu + vc_\sigma \quad (2)$$

Reducing the (\bar{w} -correlated) volatility in the recipient's number of offspring ($b_\sigma > 0$) confers on the recipient greater relative fitness in poor environmental states: extra offspring are disproportionately valuable when competitors produce few offspring²², underscoring the principle that the ultimate currency for benefits and costs under stochasticity is the expectation of relative fitness¹. It is straightforward to derive the established asocial bet-hedging model³ by setting $r=0$ (Supplementary Information A3).

¹School of Biological Sciences, University of Bristol, 24 Tyndall Avenue, Bristol BS8 1TQ, UK. ²Centre for Research in Animal Behaviour, College of Life and Environmental Sciences, University of Exeter, Exeter EX4 4QG, UK. ³Department of Genetics, Evolution and Environment, University College London, Gower Street, London WC1E 6BT, UK.

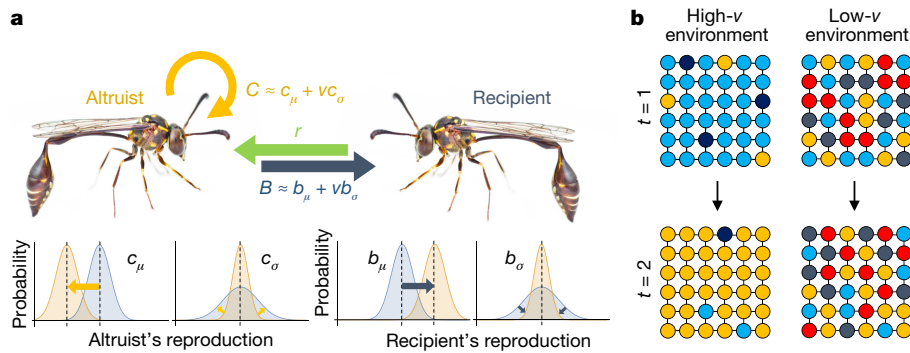


Figure 1 | Environmental stochasticity has been missing from models of social evolution. In the means-based application of Hamilton's rule ($rb_\mu > c_\mu$) to real-world organisms²¹, recipients gain an increase in average reproductive success ($b_\mu > 0$) whereas actors suffer a decrease in average reproductive success ($c_\mu > 0$). **a**, We derive an explicitly stochastic Hamilton's rule: $r(b_\mu + vb_\sigma) > c_\mu + vc_\sigma$. This shows that benefits can also arise by reducing the volatility of the reproductive success of the recipient ($b_\sigma > 0$), which depends on the magnitude of environmental stochasticity (v). An increase in the reproductive volatility of the actor ($c_\sigma > 0$) imposes a cost on the actor. Each effect represents a transformation of a probability

distribution for reproductive success (bottom). Total benefits and costs (B and C) are measured in expected relative fitness¹. **b**, Environmental stochasticity (v) is highest when spatial patches fluctuate in sync: for instance, if drought affects a randomly chosen patch Z , it should be likely that it also affects a randomly chosen patch Y (Supplementary Information A6). Here, following ref. 3, we represent patches in a lattice connected by dispersal. Colours denote environmental condition on patches at sequential time points t . See Supplementary Information A. Image of wasp reproduced with permission from Z. Soh.

Formally, we define altruistic bet-hedging as a reduction in the reproductive volatility of a recipient (positive b_σ) that overcomes an otherwise deleterious cost to the expected reproductive success of the actor (positive c_μ). Strong benefits can arise when b_μ and b_σ are both positive, and reductions in the actor's own reproductive volatility ($c_\sigma < 0$) diminish total costs (Fig. 2a, b). Moreover, when $b_\sigma > c_\sigma$, increasing stochasticity reduces the minimum relatedness (r) required for altruism to evolve (Fig. 2c). Fluctuations in relatedness (r) alter selection only if they correlate with strong fluctuations in population average reproductive success (\bar{w}) (Supplementary Information A4).

We note four predictions of the stochastic Hamilton's rule that differ from standard expectations:

(i) Selection can favour altruism ($C > 0$) with zero increase to the expected reproductive success of the recipient ($b_\mu = 0$). Such a seemingly paradoxical lack of benefits is observed in cases for which additional helpers appear redundant²³. Paradoxical helpers can be selected for by reducing the reproductive volatility of the recipient if:

$$rb_\sigma > \frac{c_\mu}{v} + c_\sigma$$

(ii) Actors may be selected to harm the expected reproductive success of their relatives ($b_\mu < 0$, $c_\mu > 0$). The harm is outweighed by a reduction in the reproductive volatility of the recipient (Fig. 2) if:

$$rb_\sigma > \frac{c_\mu - rb_\mu}{v} + c_\sigma$$

(iii) Altruists that reduce the reproductive volatility of their recipients can be favoured by selection in the absence of environmental stochasticity, but only when population size (N) is low (in extremely small populations³ or small demes with intense local competition²⁴) and $b_{\sigma^2} > c_{\sigma^2}$. Effects on variance, σ^2 , not volatility, are used here for notational convenience (Supplementary Information A5):

$$r \left(b_\mu + \frac{b_{\sigma^2}}{N \mathbb{E}_\pi[\bar{w}]} \right) > c_\mu + \frac{c_{\sigma^2}}{N \mathbb{E}_\pi[\bar{w}]} \quad (3)$$

(iv) Very strong altruistic effects ($b_\sigma \gg 0$) can undermine the success of the altruist genotype (Extended Data Fig. 1; Supplementary Information B1–B4). Altruists that substantially reduce the

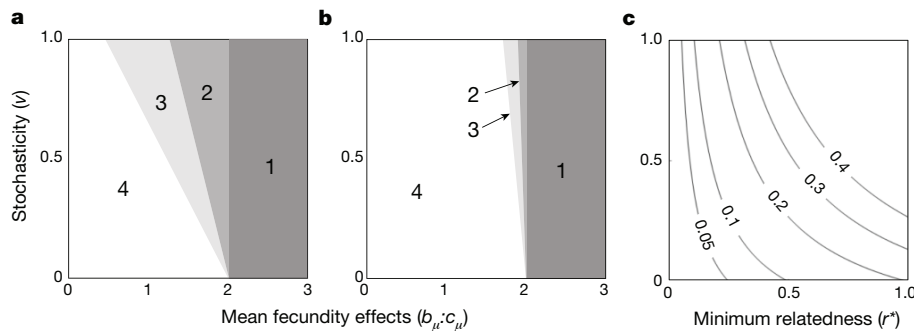


Figure 2 | Increased stochasticity can increase the potential for selection of altruistic behaviour. Without stochastic effects, altruism evolves when $rb_\mu > c_\mu$ (shown in region '1' in **a** and **b** for $c_\mu = 1$, and $r = 0.5$). As stochasticity v increases, the power of $b_\sigma:c_\sigma$ benefits increases, reducing the ratio of $b_\mu:c_\mu$ needed for the evolution of altruism. **a**, In this scenario, altruists secure a high $b_\sigma = 0.75$, considerably increasing the scope for altruism (extending region '1' to region '2'). Actors may also reduce the volatility of their personal fecundity (here, $c_\sigma = -0.4$), reducing the magnitude of the total cost C below c_μ and increasing the potential for

altruism further (extending to region '3'). Altruism is always deleterious in region '4'. **b**, In this scenario, altruists secure a low $b_\sigma = 0.1$ and personal volatility reduction of $c_\sigma = -0.1$ (regions as in **a**). Comparing **a** ($b_\sigma = 0.75$) and **b** ($b_\sigma = 0.1$), larger reductions of recipient volatility (higher b_σ) result in larger increases in the inclusive fitness of the actor. **c**, The minimum relatedness required for the evolution of altruism under different c_μ values (curved lines, from $c_\mu = 0.05$ to 0.4 , when $b_\sigma = 0.75$, $c_\sigma = 0$ and $b_\mu = 0.2$); as stochasticity (v) increases, the minimum required relatedness (r^*) decreases.

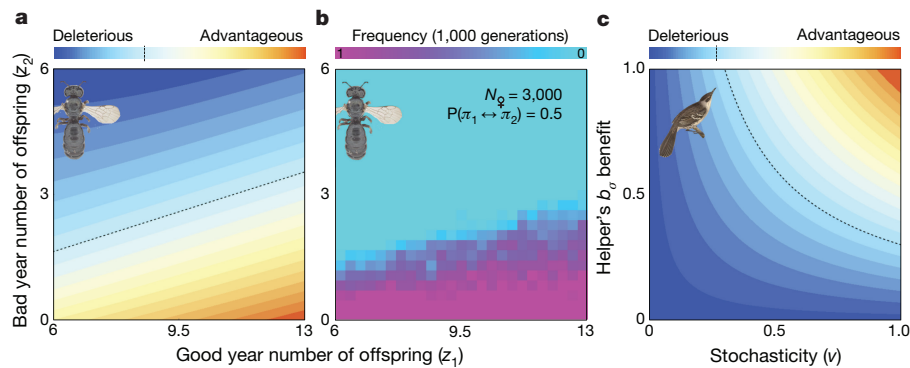


Figure 3 | Empirical studies of Hamilton's rule may benefit from incorporating stochasticity. **a**, Model of sister–sister cooperation between facultatively social insects: the means-based Hamilton's rule ($rb_{\mu} > c_{\mu}$) is violated throughout the plot. Despite this, in the region below the dashed line (which denotes $rB = C$), volatility effects can favour the invasion of nonreproductive altruists. **b**, These predictions are matched in an individual-based haplodiploid simulation. In both **a** and **b**, good and bad years occur equally ($d_{\pi} = 0.5$) at random. When benefits are slight (close to the dashed line in **a**), chance correlated fluctuations can drive cooperators extinct. In Supplementary Information B, we discuss temporal

correlation. Coordinates plot average frequency across five replicate simulations after 1,000 generations, from an initial frequency $P = 0.05$. **c**, In high-stochasticity conditions, helpers may buffer breeders from profound environmental fluctuations^{4,9,11}. We estimate rb_{μ} values in the Galapagos mockingbird, and show that volatility effects can, in principle, drive cooperation (above the dashed line) even when mean fecundity costs c_{μ} cancel out b_{μ} (here, $b_{\mu} = c_{\mu} = 0.3$). See Supplementary Information C. Image of bee, K. Walker (CC-BY 3.0 AU); image of mockingbird, Biodiversity Heritage Library (CC-BY 2.0).

reproductive volatility of their recipients spread rapidly. As successful altruists reach high frequencies, the coefficient of variation in average reproductive success ($v = \frac{\sigma_{\pi}[\bar{w}]}{\mathbb{E}_{\pi}[\bar{w}]}$) tends towards zero (Extended Data Fig. 2). When v is small, any b_{σ} has a small effect (equation (2)), so altruistic bet-hedgers undermine the condition (high v) that favoured them (Extended Data Fig. 1a, b). This frequency dependence can generate a mixed population of altruists and defectors (Extended Data Fig. 1c), provided that allele frequency does not fluctuate intensively, which can otherwise destabilize the equilibrium (Extended Data Fig. 3) and lead to fixation²⁵.

Apparent reduction of the reproductive volatility of recipients (implying $b_{\sigma} > 0$) has been shown in starlings⁹, sociable weavers²⁶, woodpeckers¹⁰, wasps²⁷ and allodapine bees¹³. We illustrate a volatility-reduction route to sociality with two examples. First, we consider sister–sister cooperation in facultatively social insects (as in certain carpenter bees, for which a means-based Hamilton's rule is violated²⁸). In strongly stochastic environments, altruism can evolve between haplodiploid sisters when values of mean fecundity alone would predict it to be deleterious, as predicted by equation (2) (Fig. 3a) and simulations of haplodiploid populations (Fig. 3b; Supplementary

Information C1). Second, using published estimates of mean fecundity and high stochasticity in Galapagos mockingbirds (*Mimus parvulus*), we indicate how volatility effects could favour cooperative breeding even if helping increases the average fecundity of the recipient only as much as it reduces that of the actor ($c_{\mu} = b_{\mu}$; Fig. 3c; Supplementary Information C2).

Equation (2) reveals three core conditions for altruistic bet-hedging. First, members of the non-altruistic genotype suffer synchronous fluctuations in lifetime reproductive success driven by environmental state (high v) that can be stabilized by sociality ($b_{\sigma} > 0$). Second, relatedness (r) is above the threshold $r^* = \frac{c_{\mu} + vc_{\sigma}}{b_{\mu} + vb_{\sigma}}$. Third, actors either cannot predict environmental fluctuations or cannot generate phenotypes for different conditions (Fig. 4; Supplementary Information B5). If actors can obtain and utilize information at sufficiently low costs (rendering the environment predictable), plastic cooperation outcompetes constitutive cooperation (increasing b_{μ} and reducing c_{μ}).

Synchronous fluctuations (high v) are generated when different patches within the population experience correlated environmental changes (Fig. 1b; Supplementary Information A6). If offspring disperse across environmentally uncorrelated patches³ but compete

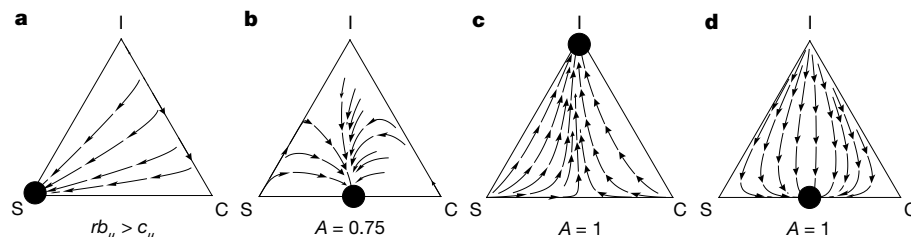


Figure 4 | The trade-off between constitutive and inducible altruism in a stochastic world depends on plasticity costs and information reliability. We show a population fluctuating randomly between a good and a bad environmental state, comprising three alleles: 'selfish' (S), for which the carriers never cooperate; 'constitutive cooperator' (C), for which the carriers always cooperate; and 'inducible cooperator' (I) for which the carriers cooperate only when they believe they are in the bad (low-fecundity) state. Information reliability is set by A (actors diagnose true state with probability A). Apexes represent monomorphic populations. Without social behaviour, individuals obtain four and one offspring in good and bad states respectively. Cooperation confers on recipients 1.5 additional offspring in bad states but reduces recipient fecundity by 0.2 offspring in good states, and costs actors 0.5 offspring in all states.

a, When considering only mean fecundity, the means-based Hamilton's rule $rb_{\mu} > c_{\mu}$, commonly used empirically, mistakenly predicts that selfishness (S) will dominate. Under stochastic conditions, cooperation evolves. **b**, Constitutive cooperators invade (until reaching a mixture of altruists and defectors) when information is imperfect ($A = 0.75$) and there is a plasticity cost (0.1 offspring). **c**, When the reliability of information is increased ($A = 1$), plastic cooperators outcompete constitutive cooperators. **d**, Increasing plasticity costs, however (here, from 0.1 to 0.3 offspring), eliminates plasticity benefits, enabling constitutive cooperators to invade. Vectors show directions of expected changes in frequencies: these represent continuous expected trajectories when frequencies are constrained to change by small amounts per generation. Relatedness $r = 0.5$ in all plots. Details are provided in Supplementary Information B.

at a whole-population level, v decreases. Likewise, iteroparity and long generations across different environmental conditions reduce v , whereas correlated exposure to environmental conditions within lifetimes increases v . For these reasons, equation (2) suggests that the most promising avenues to detect b_{σ} -driven sociality may occur among social microbes, which can experience population-wide fluctuations (high v), short generations (high v), competing clones (high r), and opportunities to confer homeostasis on others ($b_{\sigma} > 0$), including through the construction of biofilms²⁹ and incipiently-multicellular clusters withstanding profound abiotic and biotic stress.

We have shown that altruistic effects on recipient volatility are visible to selection. Notably, Hamilton's rule identifies ultimate payoffs by incorporating any effects of population structure¹. To make case-specific predictions, researchers should, accordingly, utilize explicit information on population structure and ecology. The empirical challenge to detect volatility-suppressing sociality in wild organisms will best be met using tailored models guided by field data for specific scenarios, led by the general framework of inclusive fitness theory^{1,21,30}. In summary, Hamilton's rule reveals the action of selection under stochasticity: shielding relatives from a volatile world can drive the evolution of sociality.

Code Availability Simulation output was generated using MATLAB code provided in section D of the Supplementary Information; this is also available from the corresponding author upon reasonable request.

Data Availability The data that support the findings of this study are available from the corresponding author upon reasonable request.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 July 2017; accepted 1 February 2018.

Published online 7 March 2018.

- Gardner, A., West, S. A. & Wild, G. The genetical theory of kin selection. *J. Evol. Biol.* **24**, 1020–1043 (2011).
- Hamilton, W. D. The genetical evolution of social behaviour. I. *J. Theor. Biol.* **7**, 1–16 (1964).
- Starrfelt, J. & Kokko, H. Bet-hedging—a triple trade-off between means, variances and correlations. *Biol. Rev. Camb. Phil. Soc.* **87**, 742–755 (2012).
- Cockburn, A. & Russell, A. F. Cooperative breeding: a question of climate? *Curr. Biol.* **21**, R195–R197 (2011).
- Rubenstein, D. R. & Lovette, I. J. Temporal environmental variability drives the evolution of cooperative breeding in birds. *Curr. Biol.* **17**, 1414–1419 (2007).
- Jetz, W. & Rubenstein, D. R. Environmental uncertainty and the global biogeography of cooperative breeding in birds. *Curr. Biol.* **21**, 72–78 (2011).
- Kocher, S. D. *et al.* Transitions in social complexity along elevational gradients reveal a combined impact of season length and development time on social evolution. *Proc. R. Soc. Lond. B* **281**, 20140627 (2014).
- Sheehan, M. J. *et al.* Different axes of environmental variation explain the presence vs. extent of cooperative nest founding associations in *Polistes* paper wasps. *Ecol. Lett.* **18**, 1057–1067 (2015).
- Rubenstein, D. R. Spatiotemporal environmental variation, risk aversion, and the evolution of cooperative breeding as a bet-hedging strategy. *Proc. Natl Acad. Sci. USA* **108**, 10816–10822 (2011).
- Koenig, W. D. & Walters, E. L. Temporal variability and cooperative breeding: testing the bet-hedging hypothesis in the acorn woodpecker. *Proc. R. Soc. Lond. B* **282**, 20151742 (2015).
- Shen, S., Emlen, S. T., Koenig, W. D. & Rubenstein, D. R. The ecology of cooperative breeding behaviour. *Ecol. Lett.* **20**, 708–720 (2017).
- Griesser, M., Drobniak, S. M., Nakagawa, S. & Botero, C. A. Family living sets the stage for cooperative breeding and ecological resilience in birds. *PLoS Biol.* **15**, e2000483 (2017).

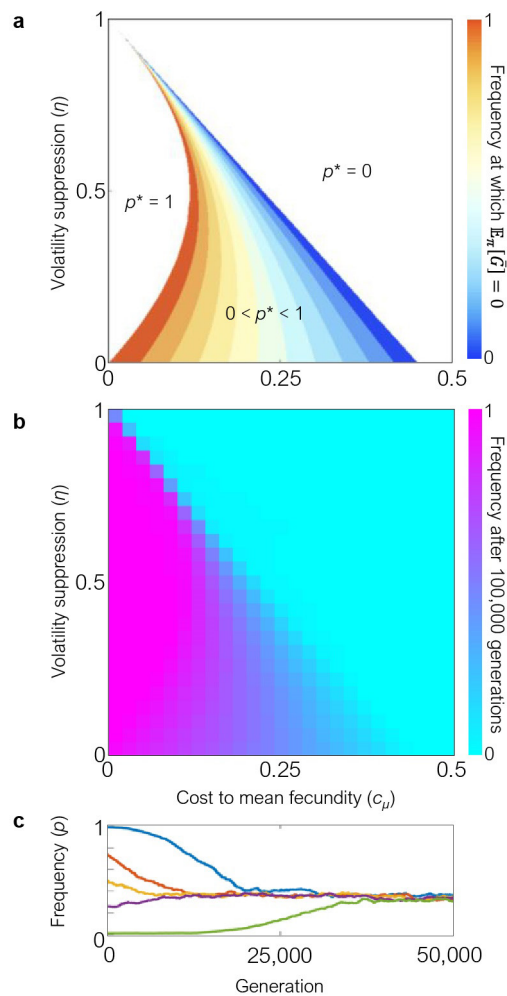
- Stevens, M. I., Hogendoorn, K. & Schwarz, M. P. Evolution of sociality by natural selection on variances in reproductive fitness: evidence from a social bee. *BMC Evol. Biol.* **7**, 153 (2007).
- Ebensperger, L. A. *et al.* Mean ecological conditions modulate the effects of group living and communal rearing on offspring production and survival. *Behav. Ecol.* **25**, 862–870 (2014).
- Sumner, S., Lucas, E., Barker, J. & Isaac, N. Radio-tagging technology reveals extreme nest-drifting behavior in a eusocial insect. *Curr. Biol.* **17**, 140–145 (2007).
- Lukas, D. & Clutton-Brock, T. Climate and the distribution of cooperative breeding in mammals. *R. Soc. Open Sci.* **4**, 160897 (2017).
- Grafen, A. Optimization of inclusive fitness. *J. Theor. Biol.* **238**, 541–563 (2006).
- Uitendhaag, J. C. M. Bet hedging based cooperation can limit kin selection and form a basis for mutualism. *J. Theor. Biol.* **280**, 76–87 (2011).
- Lehmann, L. & Rousset, F. The genetical theory of social behaviour. *Philos. Trans. R. Soc. Lond. B* **369**, 1–18 (2014).
- Koenig, W. D., Dickinson, J. L. & Emlen, S. T. in *Cooperative Breeding in Vertebrates: Studies of Ecology, Evolution, and Behavior* (eds Koenig, W. D. & Dickinson, J. L.) 353–374 (Cambridge Univ. Press, 2016).
- Bourke, A. F. G. Hamilton's rule and the causes of social evolution. *Phil. Trans. R. Soc. Lond. B* **369**, 20130362 (2014).
- Grafen, A. Developments of the Price equation and natural selection under uncertainty. *Proc. R. Soc. Lond. B* **267**, 1223–1227 (2000).
- Dickinson, J. L. & Hatchwell, B. J. in *Ecology and Evolution of Cooperative Breeding in Birds* (eds Koenig, W. D. & Dickinson, J. L.) 48–66 (Cambridge Univ. Press, 2004).
- Lehmann, L. & Balloux, F. Natural selection on fecundity variance in subdivided populations: kin selection meets bet hedging. *Genetics* **176**, 361–377 (2007).
- Lande, R. Expected relative fitness and the adaptive topography of fluctuating selection. *Evolution* **61**, 1835–1846 (2007).
- Covas, R., du Plessis, M. A. & Doutrelant, C. Helpers in colonial cooperatively breeding sociable weavers *Philetairus socius* contribute to buffer the effects of adverse breeding conditions. *Behav. Ecol. Sociobiol.* **63**, 103–112 (2008).
- Wenzel, J. W. & Pickering, J. Cooperative foraging, productivity, and the central limit theorem. *Proc. Natl Acad. Sci. USA* **88**, 36–38 (1991).
- Rehan, S. M., Richards, M. H., Adams, M. & Schwarz, M. P. The costs and benefits of sociality in a facultatively social bee. *Anim. Behav.* **97**, 77–85 (2014).
- Lowery, N. V., McNally, L., Ratcliff, W. C. & Brown, S. P. Division of labor, bet hedging, and the evolution of mixed biofilm investment strategies. *MBio* **8**, e00672-17 (2017).
- Birch, J. The inclusive fitness controversy: finding a way forward. *R. Soc. Open Sci.* **4**, 170335 (2017).

Supplementary Information is available in the online version of the paper.

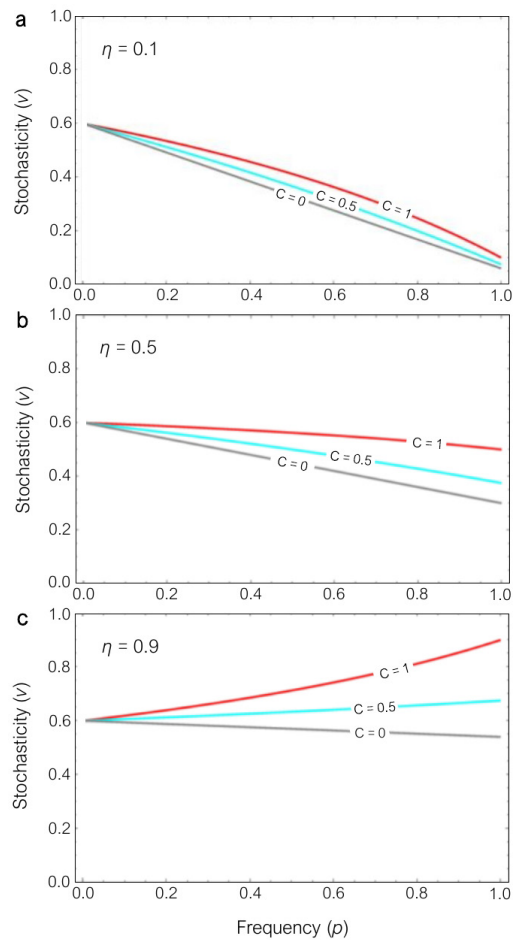
Acknowledgements We thank A. Gardner for discussions in the early stages of this work, and P.K. thanks the Behaviour Discussion Group at the Smithsonian Tropical Research Institute in Panama for the opportunity to present and discuss these ideas. We thank S. Schindler, S. Okasha, B. Autzen, J. McNamara and M. Bentley for comments on the project. P.K. was supported by the National Geographic Society (GEF-NE 145-15) and a University of Bristol Research Studentship, A.D.H. by the Natural Environment Research Council (NE/L011921/1), A.N.R. by a European Research Council Consolidator Grant (award no. 682253) and S.S. by the Natural Environment Research Council (NE/M012913/2).

Author Contributions P.K. conceived the idea, P.K. and A.D.H. performed the modelling, A.N.R. and S.S. supervised the project. All authors discussed the ideas and wrote the manuscript.

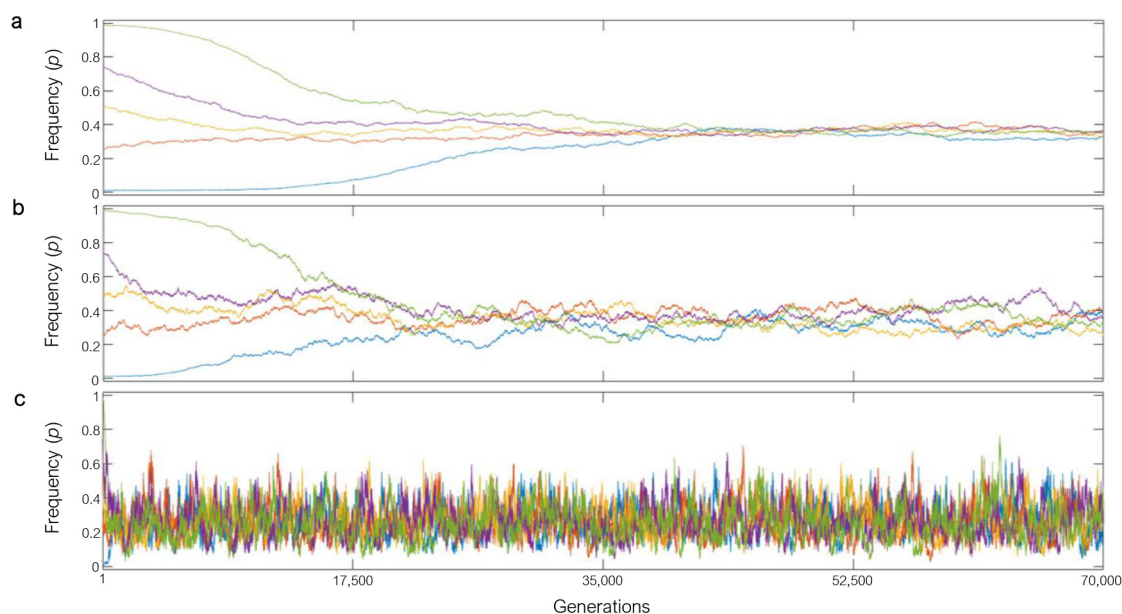
Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to P.K. (patrick.c.kennedy@live.co.uk).



Extended Data Figure 1 | The interaction between the frequency of altruists and the effectiveness of altruism. **a**, The stochastic Hamilton's rule predicts that selection on volatility-suppressing altruism with fixed costs and benefits can generate negative frequency dependence and is sensitive to mild mean-fecundity costs (c_μ). Lower values of η denote greater buffering of recipients from the environment. We evaluate a population undergoing synchronous fluctuations to identify the frequency p^* at which there is no expected change in allele frequency. We illustrate the result with individual fecundities in good years (z_1) of four offspring and in bad years (z_2) of one offspring. Relatedness is $r = 0.5$. **b**, Simulated population outcomes (frequency after 100,000 generations) match predictions of the stochastic Hamilton's rule in **a**. Warmer colours (pink) denote higher polymorphic frequencies of altruists. In this haploid model (Supplementary Information B1-4), 1% of breeding spots are available each year for replacement by offspring that year: with such constraints on the magnitude of the response to selection, radical stochastic shifts in allele frequency over single generations do not occur, allowing the population to settle at equilibria where all alleles have equal expected relative fitness without being continually displaced (Extended Data Fig. 3). **c**, Competing an altruistic allele against a defector allele reveals the action of frequency-dependent selection. Here, populations experiencing costs of $c = 0.2$ and $\eta = 0.466$ converge to $p^* = 0.359$ from any initial frequency (coloured lines show five starting frequencies from 0.001 to 0.999), as predicted by the stochastic Hamilton's rule.



Extended Data Figure 2 | Stochasticity as a function of bet-hedger frequency. Stochasticity $v = \frac{(p\eta + (1-p))\sigma_{00}}{\mu_{00} - pc}$ for the model of altruistic bet-hedging in Supplementary Information B plotted against frequency (p) and cost (c) for three different values of η . **a, b,** When η is small, representing high levels of volatility suppression, v declines steeply with p across the range of costs. **c,** When η is large, the sign of the effect of p on v depends on c . Values of other parameters: $z_1 = 4$, $z_2 = 1$, and frequency of good years $d = 0.5$.



Extended Data Figure 3 | Weak selection negates the capacity of temporal autocorrelation to drive the frequency of altruistic bet-hedgers away from the convergence frequency. Individual-based simulations from five different initial frequencies of an altruistic bet hedging allele (p) competing against a non-cooperator. **a**, The population has zero temporal autocorrelation (environmental state in each generation is random). **b**, The population has strong temporal autocorrelation (environmental state in the next generation has a 90% probability of remaining the same as in the current generation). Despite higher

amplitude fluctuations, this population converges to the same point (from the five different starting frequencies) as the uncorrelated population (**a**). **c**, The same population is simulated with greater gene frequency changes (10% of the resident genotype frequencies are available to change each generation). The population is repeatedly carried to frequencies far from the convergence point. In this case, the utility of the stochastic Hamilton's rule is both identifying whether a given trait is immune from invasion by competitors, and identifying the expected generational change at each frequency p . Parameters are $z_1 = 4$, $z_2 = 1$, $r = 0.5$.

Extended Data Table 1 | Parameters of the model

Notation	Definition	Expression
N	Population size	—
w_x	Number of surviving offspring (reproductive success) of the x th individual	—
\bar{w}	Mean reproductive success in the population	—
π	Environmental state within the set of states Π	—
G_x	Genetic value of individual x	—
r	Relatedness	β_{G_y, G_x}
z_x	Trait value of individual x	—
$\ll^k \bar{w} \gg$	k th central moment of \bar{w} across Π	$\mathbb{E}_\pi[(\bar{w} - \mathbb{E}_\pi[\bar{w}])^k]$
$\ll w_x, {}^k \bar{w} \gg$	k th mixed moment of reproductive success of individual x and \bar{w} across Π	$\mathbb{E}_\pi[(w_x - \mathbb{E}_\pi[w_x])(\bar{w} - \mathbb{E}_\pi[\bar{w}])^k]$
ν	Stochasticity of the environment	$\frac{\sigma_\pi[\bar{w}]}{\mathbb{E}_\pi[\bar{w}]}$
ρ_x	Correlation between w_x and \bar{w} across Π	$\frac{\mathbb{E}_\pi[w_x \bar{w}] - \mathbb{E}_\pi[w_x] \cdot \mathbb{E}_\pi[\bar{w}]}{\sigma_\pi[w_x] \sigma_\pi[\bar{w}]}$
B	Total benefit in Hamilton's rule under stochasticity	$\beta_{\mathbb{E}_\pi[\frac{w_x}{\bar{w}}], G_x}$ Partial regression of a focal individual's genetic value on a social partner's expected relative fitness
C	Total cost in Hamilton's rule under stochasticity	$-\beta_{\mathbb{E}_\pi[\frac{w_x}{\bar{w}}], G_x}$ Partial regression of a focal individual's genetic value on its own expected relative fitness
b_μ	Mean fecundity benefit in stochastic Hamilton's rule	$\beta_{\mathbb{E}_\pi[w_y], G_x}$ Partial regression of a focal individual's genetic value on a social partner's expected number of offspring. We make use of the identity $\beta_{\mathbb{E}_\pi[w_y], G_x} = \beta_{\mathbb{E}_\pi[w_x], G_y}$ in non-class-structured populations.
c_μ	Mean fecundity cost in stochastic Hamilton's rule	$-\beta_{\mathbb{E}_\pi[w_x], G_x}$ Partial regression of a focal individual's genetic value on its own expected number of offspring
b_σ	Volatility-suppressing benefit in stochastic Hamilton's rule	$-\beta_{\rho \sigma_\pi[w_y], G_x}$ Partial regression of a focal individual's genetic value on a partner's standard deviation in reproductive success, where the standard deviation is weighted by its correlation with \bar{w} . We make use of the identity $\beta_{\rho \sigma_\pi[w_y], G_x} = \beta_{\rho \sigma_\pi[w_x], G_y}$ in non-class-structured populations
c_σ	Volatility-suppressing cost in stochastic Hamilton's rule	$\beta_{\rho \sigma_\pi[w_x], G_x}$ Partial regression of a focal individual's genetic value on a partner's standard deviation in reproductive success, where the standard deviation is weighted by its correlation with \bar{w}
b_k	k th moment benefit in stochastic Hamilton's rule	$\beta_{\ll w_y, {}^k \bar{w} \gg, G_x}$ Partial regression of a focal individual's genetic value on the k th mixed moments of a partner's joint distribution for reproductive success w_y and population average reproductive success \bar{w} . We make use of the identity $\beta_{\ll w_x, {}^k \bar{w} \gg, G_y} = \beta_{\ll w_y, {}^k \bar{w} \gg, G_x}$ in non-class-structured populations
c_k	k th moment cost in stochastic Hamilton's rule	$-\beta_{\ll w_x, {}^k \bar{w} \gg, G_x}$ Partial regression of a focal individual's genetic value on the k th mixed moments of its own joint distribution for reproductive success w_y and population average reproductive success \bar{w} .

For derivation of regression slopes, see Supplementary Information A.

Pursuing sustainable productivity with millions of smallholder farmers

Zhenling Cui¹, Hongyan Zhang¹, Xinping Chen¹, Chaochun Zhang¹, Wenqi Ma², Chengdong Huang¹, Weifeng Zhang¹, Guohua Mi¹, Yuxin Miao¹, Xiaolin Li¹, Qiang Gao³, Jianchang Yang⁴, Zhaohui Wang⁵, Youliang Ye⁶, Shiwei Guo⁷, Jianwei Lu⁸, Jianliang Huang⁸, Shihua Lv⁹, Yixiang Sun¹⁰, Yuanying Liu¹¹, Xianlong Peng¹¹, Jun Ren¹², Shiqing Li¹³, Xiping Deng¹³, Xiaojun Shi¹⁴, Qiang Zhang¹⁵, Zhiping Yang¹⁵, Li Tang¹⁶, Changzhou Wei¹⁷, Liangliang Jia¹⁸, Jiwang Zhang¹⁹, Mingrong He¹⁹, Yanan Tong⁵, Qiuyan Tang²⁰, Xuhua Zhong²¹, Zhaohui Liu²², Ning Cao²³, Changlin Kou²⁴, Hao Ying¹, Yulong Yin¹, Xiaoqiang Jiao¹, Qingsong Zhang¹, Mingsheng Fan¹, Rongfeng Jiang¹, Fusuo Zhang¹ & Zhengxia Dou²⁵

Sustainably feeding a growing population is a grand challenge^{1–3}, and one that is particularly difficult in regions that are dominated by smallholder farming. Despite local successes^{4–8}, mobilizing vast smallholder communities with science- and evidence-based management practices to simultaneously address production and pollution problems has been infeasible. Here we report the outcome of concerted efforts in engaging millions of Chinese smallholder farmers to adopt enhanced management practices for greater yield and environmental performance. First, we conducted field trials across China's major agroecological zones to develop locally applicable recommendations using a comprehensive decision-support program. Engaging farmers to adopt those recommendations involved the collaboration of a core network of 1,152 researchers with numerous extension agents and agribusiness personnel. From 2005 to 2015, about 20.9 million farmers in 452 counties adopted enhanced management practices in fields with a total of 37.7 million cumulative hectares over the years. Average yields (maize, rice and wheat) increased by 10.8–11.5%, generating a net grain output of 33 million tonnes (Mt). At the same time, application of nitrogen decreased by 14.7–18.1%, saving 1.2 Mt of nitrogen fertilizers. The increased grain output and decreased nitrogen fertilizer use were equivalent to US\$12.2 billion. Estimated reactive nitrogen losses averaged 4.5–4.7 kg nitrogen per Megagram (Mg) with the intervention compared to 6.0–6.4 kg nitrogen per Mg without. Greenhouse gas emissions were 328 kg, 812 kg and 434 kg CO₂ equivalent per Mg of maize, rice and wheat produced, respectively, compared to 422 kg, 941 kg and 549 kg CO₂ equivalent per Mg without the intervention. On the basis of a large-scale survey (8.6 million farmer participants) and scenario analyses, we further demonstrate the potential impacts of implementing the enhanced management practices on China's food security and sustainability outlook.

Food security, environmental degradation and climate change are grand challenges facing humankind^{1,2}. Agriculture is at the heart of these challenges, as food production must be increased by 60–110% (from 2005) to meet the growing demand by 2050^{3,9}, and at the same

time adverse environmental impacts need to be reduced amid climate change and growing competition for natural resources^{10,11}. The greatest challenge occurs in regions in which smallholder farming dominates the agricultural landscape, for example, in sub-Saharan Africa, India and China. In these regions, food security and sustainability depend on how smallholders, who are typically resource-limited and knowledge-poor, farm their land¹². Much effort has endeavoured to enhance smallholder productivity^{13–15}. However, mobilizing millions of smallholder farmers and encouraging them to adopt management technologies that simultaneously address production and pollution problems has been infeasible. The need to do so is particularly important in countries in which smallholders operate high-input, low-efficiency systems.

China is a case in point. With 200–300 million households that each farm a few hectares of land, the agricultural system relies heavily on high-to-excessive inputs. For example, nitrogen application averages to 305 kg N ha⁻¹ yr⁻¹ compared to 74 kg N ha⁻¹ yr⁻¹ worldwide¹⁶; nitrogen use efficiency (the fraction of nitrogen input harvested as product) is only 0.25 compared to 0.42 worldwide and 0.65 in North America¹⁷. Over-application of nitrogen has caused widespread soil acidification¹⁸, devastating water pollution¹⁹ and excessive greenhouse gas (GHG) emissions²⁰. For a sustainable food-secure future, China needs a 'great balancing act'²¹ to attain high yield and high efficiency with a substantially reduced environmental footprint. This cannot be achieved without the vast smallholder-farming communities.

Here we report the outcome of nationally coordinated efforts over a 10-year period that encouraged 20.9 million smallholders to adopt enhanced management technologies for greater yield and reduced environmental pollution. First we present the results of 13,123 field trials that tested the applicability of a comprehensive decision-support integrated soil–crop system management (ISSM) program for growing maize, rice and wheat across China's vast agroecological zones. We then describe coordinated campaigns, leading to the implementation of ISSM-based management in farmland with a total of 37.7 million cumulative hectares over the years (2005–2015). Finally, we discuss

¹College of Resources and Environmental Sciences, China Agricultural University, Beijing 100193, China. ²College of Resources and Environmental Sciences, Agricultural University of Hebei, Baoding 071001, China. ³College of Resources and Environment, Jilin Agricultural University, Changchun 130118, China. ⁴Key Laboratory of Crop Genetics and Physiology of Jiangsu Province, Yangzhou University, Yangzhou 225009, China. ⁵Northwest Agriculture and Forestry University, Yangling 712100, China. ⁶College of Resources and Environmental Sciences, Henan Agricultural University, Zhengzhou 450000, China. ⁷College of Resources and Environmental Sciences, Nanjing Agricultural University, Nanjing 210095, China. ⁸Huazhong Agricultural University, Wuhan 430070, China. ⁹Soil and Fertilizer Research Institute, Sichuan Academy of Agricultural Sciences, Chengdu 610066, China. ¹⁰Soil and Fertilizer Research Institute, Anhui Academy of Agricultural Sciences, Hefei 230031, China. ¹¹College of Resources and Environment, Northeast Agricultural University, Harbin 150030, China. ¹²Institute of Agricultural Resource and Environment, Jilin Academy of Agricultural Sciences, Changchun 130033, China. ¹³State Key Laboratory of Soil Erosion and Dryland Farming on the Loess Plateau, Northwest Agriculture and Forestry University, Yangling 712100, China. ¹⁴College of Resources and Environment, Southwest University, Chongqing 400716, China. ¹⁵Institute of Agricultural Environment and Resources, Shanxi Academy of Agricultural Sciences, Taiyuan 030031, China. ¹⁶College of Resources and Environmental Science, Yunnan Agricultural University, Kunming 650201, China. ¹⁷Department of Resources and Environment, Agriculture College, Shihezi University, Shihezi, Xinjiang, 832003, China. ¹⁸Institute of Agricultural Resources and Environment, Hebei Academy of Agriculture and Forestry Sciences, Shijiazhuang 050051, China. ¹⁹College of Agronomy, Shandong Agricultural University, Tai'an 271000, China. ²⁰College of Agronomy, Hunan Agricultural University, Changsha 410128, China. ²¹Rice Research Institute, Guangdong Academy of Agricultural Science, Guangzhou, Guangdong 510640, China. ²²Shandong Academy of Agricultural Sciences, Jinan 250100, China. ²³College of Plant Science, Jilin University, Changchun, 130012, China. ²⁴Institute of Plant Nutrition and Resource Environment, Henan Academy of Agricultural Sciences, Zhengzhou 450002, China. ²⁵Center for Animal Health and Productivity, School of Veterinary Medicine, University of Pennsylvania, Kennett Square, Pennsylvania 19348, USA.

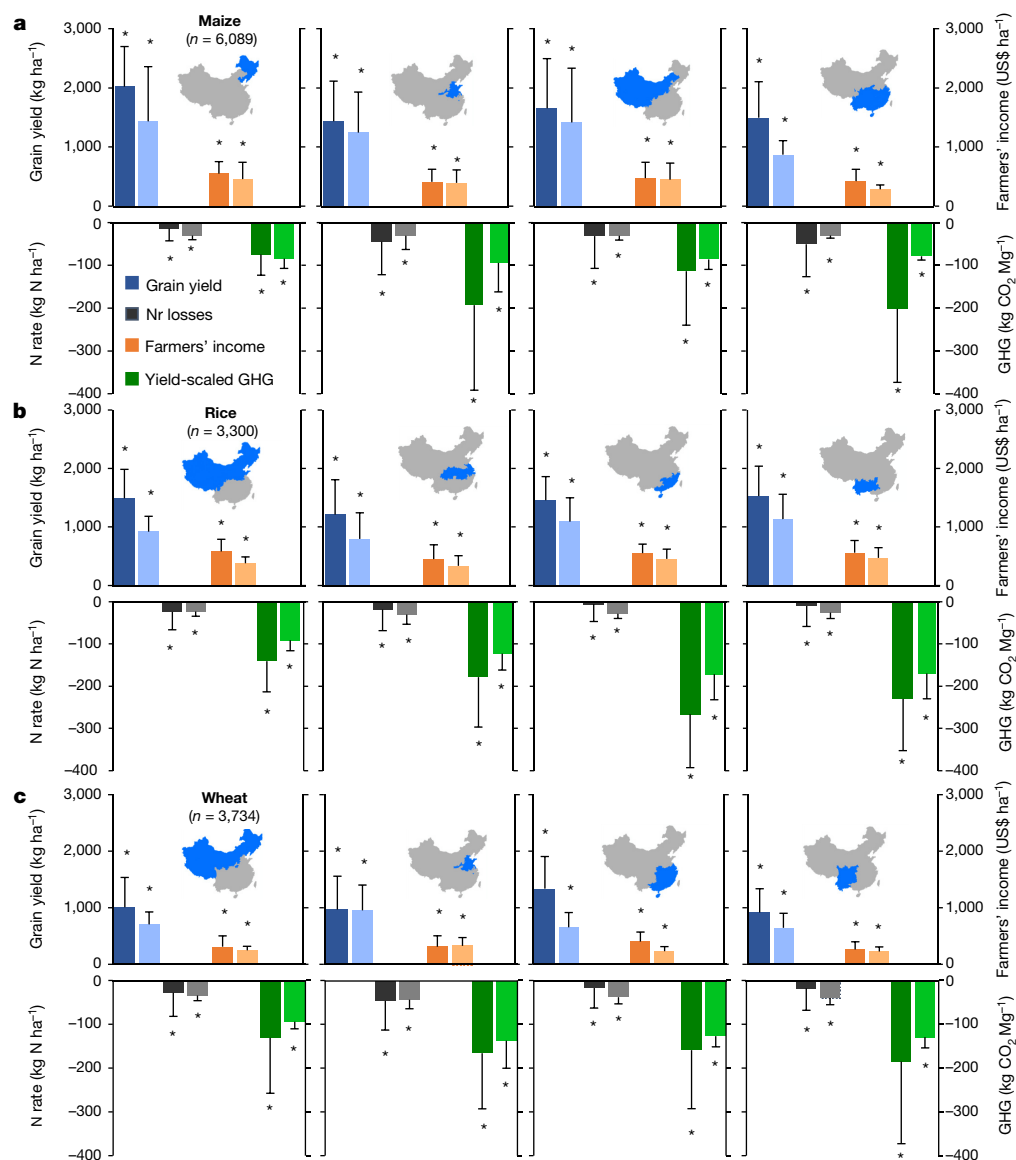


Figure 1 | Production and environmental performance with ISSM-based intervention. **a–c**, Yield response, nitrogen rate, farmers' net income and GHG emissions of ISSM-based management compared to farmers' practices (0 baseline) for maize (**a**), rice (**b**) and wheat (**c**) in four agroecological zones in China. Dark-coloured bars denote data from field trials ($n = 13,123$),

light-coloured bars indicate data from the national campaign during 2005–2015. Nr, reactive nitrogen. Data are mean \pm s.d. * $P < 0.05$, significant difference between treatment (ISSM) and control (farmers' practice). Chinese map was obtained from the Resource and Environment Data Cloud Platform (<http://www.resdc.cn/data.aspx?DATAID=202>).

scenarios for pursuing sustainable productivity in the entire country and the potential impacts on grain output and selected environmental indices.

To start, we needed technological tools for designing management practices that can be packaged for making field recommendations. Such technologies need to be comprehensive and include key crop–soil–water–nutrient parameters, as well as adaptive, in order to suit different biophysical conditions. The ISSM framework^{4,22} appears to suit these needs. It consists of a crop module from which cropping strategies (for example, crop variety, planting date and density for maize, rice or wheat) can be determined based on crop model simulations for optimal use of solar and thermal resources in a given region; and a resource supply module for the formulation of nutrient and water applications according to soil tests and the needs of the growing crops. Previous studies have demonstrated that following ISSM-based recommendations resulted in greater yields (18–35%), a reduction in nitrogen fertilizer usage (4–14%) and improved nitrogen productivity (kg grain produced per kg nitrogen applied; 32–46%) compared to the conventional practices of the farmers⁵.

We investigated whether ISSM could be used across China's major agroecological zones, which range from frigid to subtropical, and from arid to semi-arid to humid, to obtain similar outcomes across all regions. We conducted field trials, with a total of 13,123 site years between 2005 and 2015 across agroecological zones (Extended Data Fig. 1). Each trial included ISSM-based recommendations (treatment) compared to the conventional practice of the farmers (control), with the participating farmer carrying out field operations and campaign collaborators providing on-site guidance (Methods and Extended Data Table 1). Yield response to treatment varied for different crops in different agroecological zones. But in all cases ISSM-based treatment enhanced yield, nitrogen productivity, and farmer profitability and reduced nitrogen losses. Averaged over all site years, grain yields increased from 7.83 to 9.54 Mg ha⁻¹ for maize ($n = 6,089$), from 7.03 to 8.41 Mg ha⁻¹ for rice ($n = 3,300$) and from 5.69 to 6.73 Mg ha⁻¹ for wheat ($n = 3,734$). At the same time, nitrogen rate (amount of nitrogen applied per unit area (kg N ha⁻¹)) decreased by 8.5–15.6% (Fig. 1). As expected, ISSM-based treatment led to greater nitrogen use efficiency, net income and environmental performance. Nitrogen productivity

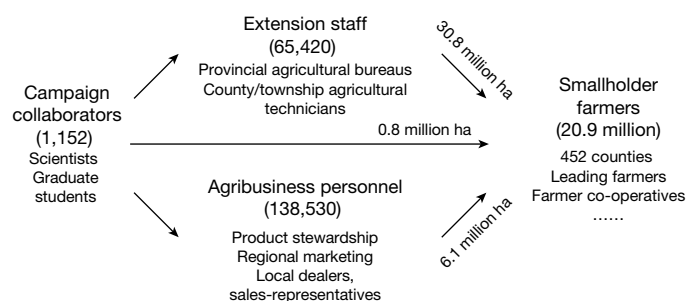


Figure 2 | The national campaign network. The campaigns were conducted to encourage smallholder farmers to implement ISSM-based management practices for high yield, high efficiency and low pollution. Campaign collaborators (scientists and graduate students at agricultural universities or research institutions) developed locally applicable ISSM-based recommendations and provided training to extension staff (agricultural technicians and field agents at various governmental agencies) and agribusiness personnel (agricultural production supply and services, including headquarters business/product development, regional marketing, and local dealers and sales representatives). All three entities worked with farmers. Numbers in parentheses indicate the number of individuals involved in the campaign.

increased by 26.0–33.1%. Calculated reactive nitrogen losses (Methods) decreased by 22.9–34.9% and GHG emissions were reduced by 18.6–29.1% (Fig. 1).

The expansive field trials provided strong evidence that the ISSM program is robust and versatile and could be used nationwide for developing management practices to simultaneously enhance productivity and environmental performance. Once ISSM-based recommendations were derived from the field trials of a given region, we led and coordinated campaign activities to promote their adoption throughout the region (Methods). The national campaign consisted of more than 1,000 collaborators, 65,000 extension agents and 130,000 agribusiness personnel (Fig. 2), who engaged 20.9 million farmers in 452 counties to implement ISSM-based practices in fields with a total of 37.7 million cumulative hectares over the years (2005–2015).

Production and environmental outcomes from the national campaign were in line with expectations. Aggregated 10-year data showed an overall yield improvement of 10.8–11.5% and a reduction in the use of nitrogen fertilizers of 14.7–18.1%, when comparing ISSM-based interventions and the prevailing practices of the farmers (Table 1 and Methods). This led to a net increase of 33 Mt grains and a decrease of 1.2 Mt nitrogen fertilizer use during the 10-year period, equivalent to US\$12.2 billion (Table 1). To put the numbers in perspective, Malawi's total grain output was 31.5 Mt during 2005–2014²³; nitrogen fertilizer use in the entire sub-Saharan Africa was 4.6 Mt during 2005–2015²³.

We assessed relevant environmental impacts by calculating reactive nitrogen losses (N_2O emission, NH_3 volatilization, NO_3^- leaching and nitrogen runoff losses) and GHG emissions (Methods). Results varied widely, depending on crop type, nitrogen rate, biophysical conditions and other factors. Aggregated results showed that ISSM-based interventions reduced reactive nitrogen losses by 13.3–21.9% and GHG emissions by 4.6–13.2%. The yield-scaled nitrogen footprint averaged 4.6, 4.7 and 4.5 kg reactive nitrogen loss per Mg of maize, rice and wheat produced, respectively, compared to 6.1, 6.0 and 6.4 kg Mg^{-1} without intervention. Similarly, yield-scaled GHG emissions were 328 kg, 812 kg and 434 kg compared to 422 kg, 941 kg and 549 kg CO_2 equivalent per Mg for maize, rice and wheat, respectively (Table 1).

Changing farmer behaviour requires more than scientifically sound and evidence-based technologies^{24,25}. Building trust, participatory innovation, developing human capacity and strengthening the coherence of the farming communities are critical for sustainable changes; we pursued these goals vigorously throughout the campaign (Methods). Examples include providing basic knowledge to progressive farmers and increasing their problem-solving skills, enabling these farmers

Table 1 | Comparison of grain output, nitrogen fertilizer use, nitrogen productivity, reactive nitrogen losses, GHG emissions and net economic gains

		Maize	Rice	Wheat	Total
Area (million ha)		12.8	17.0	7.9	37.7
Grain output (Mt)	FP	108	133	51	292
	ISSM	120	147	56	324
	Difference	11.5%	11.1%	10.8%	11.2%
N fertilizer use (Mt)	FP	2.9	3.3	1.8	8.0
	ISSM	2.5	2.8	1.5	6.8
	Difference	–14.7%	–15.1%	–18.1%	–15.6%
N productivity (kg N per kg grain)	FP	40.0	41.9	28.4	NA
	ISSM	53.4	55.1	38.5	NA
	Difference	33.4%	31.5%	35.7%	NA
Nr losses* (Mt)	FP	0.65	0.79	0.33	1.8
	ISSM	0.55	0.69	0.25	1.5
	Difference	–15.0%	–13.3%	–21.9%	–15.5%
CO_2 -equivalent emission† (Mt)	FP	45	125	28	198
	ISSM	39	119	24	183
	Difference	–12.9%	–4.6%	–13.2%	–7.7%
Net economic gain‡ (billion US\$)	FP	18.0	28.0	7.2	53.2
	ISSM	22.1	34.2	9.2	65.5
	Difference	22.7%	22.1%	27.2%	23.0%

Comparison between ISSM-based management technologies (ISSM) and conventional practices of the farmers (FP) for maize, rice and wheat during the national campaign from 2006 to 2015. Difference indicates the percentage increase of ISSM-based recommendations over the conventional practice.

*Reactive nitrogen losses include NH_3 volatilization, NO_3^- leaching, N_2O emissions and nitrogen runoff. See Methods for calculations.

†GHG emissions include CO_2 , CH_4 and N_2O from the whole life cycle of crop production. See Methods for calculations.

‡Net economic gain from increased yield and decreased nitrogen fertilizer use, calculated as 0.31, 0.40, 0.32 and 0.62 US\$ per kg of maize, rice, wheat and nitrogen, respectively.

to lead fellow villagers²⁶; fostering farmer cooperatives to give smallholders a collective voice for negotiating purchases or marketing their products at better price as well as influencing local agricultural policies (Supplementary Information). Campaign collaboration, engagement mechanisms, socioeconomic factors and relevant impacts are described in the Methods.

For China's vast numbers of smallholder farmers, we wanted to understand how varied their productivity and environmental performances were. We therefore extracted results from a large-scale survey (8.6 million participants from 1,944 counties covering 73% of total acreage of the three crops; Methods). Our analysis indicates that the majority of smallholder farmers (61%) had yields at least 10% (up to 50%) below the ISSM-based yields, while their nitrogen rates were comparable to or higher than ISSM-based rates. County-level performance scores show considerable gaps in most cases, comparing county-average yield and nitrogen rate with ISSM-based benchmarks (Fig. 3), indicating that there is room for improvement.

We then conducted scenario analyses to assess the potential impacts if all surveyed counties were to adopt ISSM-based technologies. Counties in each agroecological zone were categorized as 'low yield and high nitrogen rate', 'low yield and low nitrogen rate', 'high yield and high nitrogen rate' and 'high yield and low nitrogen rate' (Extended Data Tables 2–4). Scenario 1 targets the low yield and high nitrogen rate group with the ISSM-based yield and nitrogen rate as benchmark; scenario 2 included the low yield and high nitrogen and the low yield and low nitrogen group; and scenario 3 further added the high yield and high nitrogen rate group. Our analysis shows that compared to business as usual (that is, prevailing practices without intervention)²⁷, implementing ISSM-based technologies would increase annual grain output by 19.3, 70.5 and 82.4 Mt for scenarios 1, 2 and 3, respectively. Taken together (that is, including scenario 3), there would be an annual reduction of nitrogen fertilizer use by 1.10 Mt (8.5% compared to business as usual), reactive nitrogen losses by 0.45 Mt (16.0%) and CO_2 -equivalent emissions by 23.4 Mt (7.6%; Extended Data Table 5). The feasibility for a nationwide scale-up is discussed in

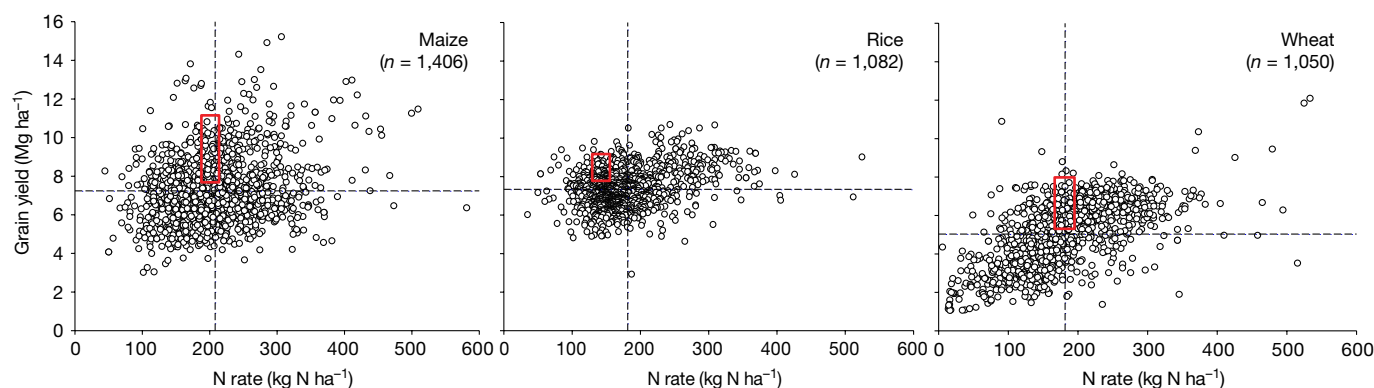


Figure 3 | National performance scores based on surveys of farmers during 2005–2014 for maize, rice and wheat. County-level performance scores (yield and nitrogen application rate) based on survey results of 8.6 million farmers in 1,944 counties during 2005–2014. **a**, Maize. **b**, Rice. **c**, Wheat. Dashed lines denote means of the total in yields and

nitrogen rates, 7.2 Mg ha⁻¹ and 208 kg N ha⁻¹ for maize, 7.3 Mg ha⁻¹ and 181 kg N ha⁻¹ for rice, 5.0 Mg ha⁻¹ and 181 kg N ha⁻¹ for wheat, respectively. Rectangular boxes indicate ranges of ISSM-based performance. Numbers in parentheses indicate the number of counties.

the Supplementary Information, along with potential limitations and possible barriers.

Worldwide, 2.5 billion smallholders farm 60% of the world's arable land²⁸. How they perform directly determines their own livelihood, and at the same time these farmers collectively impact the global food, resources and ecosystem health as a whole. Empowering smallholder farmers with enhanced management technologies to help them attain greater productivity and environmental performance is critical as we pursue an equitable world with a sustainable future. Towards this end, this study can be a valuable addition to the range of viable solutions.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 March 2017; accepted 19 January 2018.

Published online 7 March 2018.

1. Foley, J. A. *et al.* Solutions for a cultivated planet. *Nature* **478**, 337–342 (2011).
2. Matson, P. A., Parton, W. J., Power, A. G. & Swift, M. J. Agricultural intensification and ecosystem properties. *Science* **277**, 504–509 (1997).
3. Tilman, D., Balzer, C., Hill, J. & Befort, B. L. Global food demand and the sustainable intensification of agriculture. *Proc. Natl Acad. Sci. USA* **108**, 20260–20264 (2011).
4. Chen, X. P. *et al.* Integrated soil–crop system management for food security. *Proc. Natl Acad. Sci. USA* **108**, 6399–6404 (2011).
5. Chen, X. *et al.* Producing more grain with lower environmental costs. *Nature* **514**, 486–489 (2014).
6. Cui, Z. *et al.* Closing the yield gap could reduce projected greenhouse gas emissions: a case study of maize production in China. *Glob. Change Biol.* **19**, 2467–2477 (2013).
7. Ladha, J. K. *et al.* Agronomic improvements can make future cereal systems in South Asia far more productive and result in a lower environmental footprint. *Glob. Change Biol.* **22**, 1054–1074 (2016).
8. Liu, Y. *et al.* Net global warming potential and greenhouse gas intensity from the double rice system with integrated soil–crop system management: a three-year field study. *Atmos. Environ.* **116**, 92–101 (2015).
9. Alexandratos, N. How to feed the world in 2050. In *The Proceedings of a Technical Meeting of Experts* 1–32 (FAO, 2009).
10. Garnett, T. *et al.* Sustainable intensification in agriculture: premises and policies. *Science* **341**, 33–34 (2013).
11. Gunton, R. M., Firbank, L. G., Inman, A. & Winter, D. M. How scalable is sustainable intensification? *Nat. Plants* **2**, 16065 (2016).
12. Sanchez, P. A. En route to plentiful food production in Africa. *Nat. Plants* **1**, 14014 (2015).
13. Dixon, J., Gulliver, A. & Gibbon, D. *Farming Systems and Poverty* (FAO, 2001).
14. Pretty, J. Agricultural sustainability: concepts, principles and evidence. *Phil. Trans. R. Soc. B* **363**, 447–465 (2008).
15. Denning, G. *et al.* Input subsidies to improve smallholder maize productivity in Malawi: toward an african green revolution. *PLoS Biol.* **7**, e1000023 (2009).
16. IFA. IFADATA <http://ifadata.fertilizer.org/ucSearch.aspx> (International Fertilizer Industry Association, 2011).

17. Zhang, X. *et al.* Managing nitrogen for sustainable development. *Nature* **528**, 51–59 (2015).
18. Guo, J. H. *et al.* Significant acidification in major Chinese croplands. *Science* **327**, 1008–1010 (2010).
19. Diaz, R. J. & Rosenberg, R. Spreading dead zones and consequences for marine ecosystems. *Science* **321**, 926–929 (2008).
20. Zhang, W. *et al.* Closing yield gaps in China by empowering smallholder farmers. *Nature* **537**, 671–674 (2016).
21. Searchinger, T. *et al.* The great balancing act. Installment 1 of creating a sustainable food future. *World Resources Institute* <http://www.worldresourcesreport.org> (2013).
22. Zhang, F., Chen, X. & Vitousek, P. Chinese agriculture: an experiment for the world. *Nature* **497**, 33–35 (2013).
23. FAO. FAOSTAT. *Statistics Division of the Food and Agriculture Organization of the United Nations* <http://www.fao.org/faostat/en/#home> (FAO, 2016).
24. Spielman, D. J., Byerlee, D., Alemu, D. & Kelemework, D. Policies to promote cereal intensification in Ethiopia: the search for appropriate public and private roles. *Food Policy* **35**, 185–194 (2010).
25. Matson, P., Clark, W. C. & Andersson, K. *Pursuing Sustainability: a Guide to the Science and Practice* (Princeton Univ. Press, 2016).
26. Zhao, P. F. *et al.* Training and organization programs increases maize yield and nitrogen-use efficiency in smallholder agriculture in China. *Agron. J.* **108**, 1944–1950 (2016).
27. National Bureau of Statistics of China. *China Statistical Yearbook* (China Statistics Press, 2016).
28. IFAD & UNEP. *Smallholders, Food Security and the Environment* (International Fund for Agricultural Development, 2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We acknowledge all those who provided local assistance or technical help during the national campaign. We also thank J. D. Toth at the University of Pennsylvania for editing assistance. This work was financially supported by the Chinese National Basic Research Program (2015CB150400), the Innovative Group Grant from the NSFC (31421092), the Special Fund for Agro-scientific Research in the Public Interest (201103003), and National Natural Science Foundation—Outstanding Youth Foundation (31522050).

Author Contributions F.Z., X.C. and Z.C. designed the research and F.Z. supervised the project. Z.C., H.Z., G.M., Y.M., X.L., W.M., Q.G., J.Y., Z.W., Y. Ye, S.G., J.L., J.H., S. Lv, Y.S., Y.L., X.P., J.R., S. Li, X.D., X.S., Qia.Z., Z.Y., L.T., C.W., L.J., J.Z., M.H., Y.T., Q.T., X.Z., Z.L., N.C., C.K. and M.F. were key players as regional coordinators or group leaders for the field trials and national campaign; W.M., C.H., C.Z., W.Z., H.Y., Y. Yin, R.J., X.J. and Qin.Z. collected and analysed the data. Z.C., F.Z. and Z.D. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to F.Z. (zhangfs@cau.edu.cn).

Reviewer Information *Nature* thanks N. Mueller, D. Powlson, J. Reganold and L. Samberg for their contribution to the peer review of this work.

METHODS

The scope of the work reported here consists of three components: (i) field trials conducted across major agroecological zones to develop locally applicable ISSM-based recommendations; (ii) a 10-year national campaign to promote wide adoption of the enhanced management practices; and (iii) results extracted from a large-scale survey to examine, through scenario analyses, the potential impacts on grain output and environmental footprint of implementing ISSM-based technologies in the entire production systems of rice, wheat and maize in China.

The agroecological zones for maize, rice and wheat. Cereal crops are cultivated across China from frigid to subtropical and from arid to semi-arid and humid regions. On the basis of climatic conditions, geographical location and cropping systems (for example, crop type, rotation, rainfed or irrigation), we categorized the production systems into four agroecological zones for each crop. These are northeast, central, northwest and south China for maize; north, Yangtze River Basin, southeast and southwest for rice; and north, central, Yangtze River Basin and southwest for wheat. More details regarding the climate and cropping system in these zones are shown in Extended Data Fig. 1 and described in the Supplementary Information.

Field trials. A total of 13,123 site years of field trials were conducted from 2005 to 2015 for the three crops ($n = 6,089$ for maize, 3,300 for rice and 3,734 for wheat), with sites spread across all agroecological zones (Extended Data Fig. 1). A network of collaborators chose relevant locations/sites then solicited farmer participants based on willingness, field size, labour availability, land tenure, and so on. Each field trial included two types of management: conventional farmers' practice (control) and ISSM-based recommendations (treatment; developed specifically for a given area). The recommended practices were discussed with local experts and participating farmers. Adjustments were made when necessary. Finally, the agreed-upon management technologies were implemented in the fields by the farmer; the collaborators provided guidance on-site during key operations, such as sowing, fertilization, irrigation and harvest. Campaign collaborators recorded fertilizer rate, pesticide and energy use, and calculated nutrient application rate. At maturity, grain yield and aboveground biomass were sampled by the collaborators for plots with a size of 6 m^2 for wheat and rice, and 10 m^2 for maize. Plant samples were dried at 70°C in a forced-draft oven to constant weight, and grain yield was standardized at 14% moisture for all crops.

National campaign. The campaign was initiated as a national 'high yield high efficiency' (double high) umbrella project, funded via several grants (see Acknowledgements) by government agencies, for example, the Ministry of Agriculture and the Ministry of Science and Technology. The campaign was led by a group of scientists at the China Agricultural University. The core network consisted of 1,152 scientists and graduate students from 33 agricultural universities or research academies, typically at the provincial or regional level. They developed ISSM-based recommendations through field trials (see 'Field trials'), then trained extension agents ($n = 65,420$; mostly county or township agricultural technicians supported by the government) as well as private sector personnel (seed and fertilizer sales representatives, $n = 44,580$; dealers, $n = 93,950$). The extension agents received basic training from campaign collaborators and worked with farmers to promote the adoption of ISSM-based technologies. The private sector personnel participated mainly by providing farmers with the needed supplies such as specific maize cultivars or fertilizer blends based on ISSM recommendations.

Outreach activities. A variety of methods were used to disseminate ISSM-based recommendations to the farming communities. Main mechanisms included the following: (i) workshops to discuss details of ISSM-based recommendations, with (already) participating farmers sharing their experience and outcomes; (ii) on-site guidance was provided in a timely manner when needed; (iii) high-quality production materials, for example, seeds, fertilizers and other agricultural chemicals, were uniformly supplied to some sites; (iv) field day and harvest time meetings were organized to demonstrate the outcomes of the advanced management technologies; (v) ISSM-based recommendations were printed and distributed free to extension personnel and farming households, such as leaflets and customized calendars. Industry investments in advertising also communicated information about key products and practices. During the campaign, about 14,000 training workshops, 21,000 field days, and more than 6,000 site demonstrations were organized by campaign staff; more than 337,000 pamphlets were distributed.

Engaging farmers and changing behaviour. Our experience and approaches in persuading farmers to change their conventional practices can be summarized into the following aspects. First, participatory innovation is an essential step to initiate changes at any a given site. Scientifically sound and evidence-based management practices must be effectively communicated to farmers; often some modifications were made to address the specific needs of the local communities. Participatory innovation is attained through dialogues and close interactions between campaign collaborators and farmers, frequently involving local extension agents as well. These processes have been described in a previous study²⁰.

Second, enabling leading farmers to lead, the followers will follow. Lead farmers are those who have good farming skills and are open-minded to new ideas. They are recognized by fellow villagers as the smart and successful individuals whose actions inspire others. Lead farmers are active players in the participatory innovation processes during our campaign, providing inputs and feedbacks. They were the early adopters of the advanced management technologies. They influenced others by example, and helped during field days or site demonstrations by answering questions and explaining to fellow farmers what practices they adopted, why, and the attained or expected benefits. This is important, because sometimes information-driven outreach activities organized by campaign personnel, for example, training workshops or distribution of printed materials, may have limited impact on some farmers who were simply uninterested in learning the information *per se*. However, these farmers were willing to follow the steps of successful (lead) farmers. Third, changing behaviour requires building trust, which takes time and effort. For example, when starting at a new site/location, we were often asked by farmers whether we were trying to sell them something (seeds or fertilizers). Once we gained a solid foothold with demonstrated greater yield and less fertilizer use, typically in the fields of leading farmers, and with no hidden agenda, they became willing participants.

It is worth noting that farmers are not blindly following the recommendations of the scientists. Instead, they do perform, in their own way, the act of balancing potential benefits against risks as well as conforming to farming reality (for example, many farmers are only temporarily returning from their cash-earning jobs in the city to carry out tasks of agricultural operations). A recent publication²⁰ includes detailed information on specific concerns related to farmers' risk-aversion; certain compromises had to be made between farmers and the researchers involving plant density, fertilization frequency (number of split applications) and nutrient sourcing (inorganic versus organic fertilizers).

During the campaign, we also encountered barriers and experienced challenges. For example, we observed that some farmers appeared indifferent during some outreach events. We later learned that it was mainly, because they could not comprehend the scientific content that we were trying to deliver. We solved the problem by having local (county or township) agents acting as an on-site 'interpreter' in ways that speaks/connects with those farmers. Furthermore, not all recommended practices were uniformly adopted by all participating farmers. One particular challenge was rural labour shortage, because those that are young and able-bodied have gone to take city jobs²⁹. This made some of the recommended best nutrient management practices (for example, in-season fertilizer applications) difficult to implement. It is also worth noting that the interests of agribusinesses do not always align with those of our campaign staff. For example, one of our main strategies used in the campaign was to select a site (for example, a village) for a given area, establish the base with field demonstrations of ISSM-based practices, then attract and engage more farmers from the same as well as neighbouring villages, creating a snowballing and lasting effect. But sometimes, our partners in the private sector were more interested in changing sites so as to reach more farmer-clients. Vigorous debates and discussion ensued. Eventually, the private sector personnel conformed to our reasoned schemes while using the established sites as demonstrations for visitors from other areas. Notably, we have not encountered nor received negative feedbacks regarding risks projected onto farmers due to agribusinesses involvement in the high yield high efficiency project. Farmers are not obliged to purchase seeds or other production inputs from designated suppliers, although oftentimes they opt to do so for the benefit of group discount. Furthermore, those suppliers are typically large and reputable enterprises that are interested in doing long-term business.

Data collection. Farmers conducted all field operations. Campaign collaborators and/or extension agents were responsible for information and data collection. Typically, 10–30 farmers were randomly selected per ISSM-adopting site; another group of randomly selected 10–30 farmers from a nearby village without ISSM intervention served as a control/comparison. From the selected pool of farmers (roughly 14,600 paired data points), information on key management practices were obtained through a questionnaire survey, including crop varieties, planting densities, planting dates, fertilizer rates and harvest dates. For some sites, grain yields were directly measured in the same way as the field trials (see 'Field trials') for the selected 10–30 farmers. Yield and nitrogen rate were then averaged for each site.

Campaign cost. Direct costs, including the numerous field trials for developing locally applicable ISSM-based recommendations, approximated 350 million RMB in total (equivalent to US\$54 million), funded through various grants (see Acknowledgements). We do not have data on indirect expenditure through, for example, local governments that cost-shared with farmer groups that needed to purchase necessary equipment or agribusinesses that sponsored outreach activities²⁰. Direct profit, calculated from increased grain output and reduced nitrogen fertilizer use, was US\$12.2 billion (Table 1), which does not include relevant environmental benefits associated with reductions in reactive nitrogen

losses and in GHG emissions. On the basis of the rough estimates, the cost:benefit ratio would be 1:226. Note that campaign expenditure was primarily operational. In some cases, participating farmers were paid a small fee for their services. Campaign collaborators, extension staff and agribusiness personnel engaged in the campaign were not compensated for their time or efforts. Their participation in and contribution to the achievement of the campaign stem from a combination of professional duty (it is their job) and personal enthusiasm, as the campaign provided a platform for inspired individuals who desired to make a difference. This is particularly so with extension staff, who were reinvigorated through their participation in and contribution to the campaign, coupled with purposeful professional engagement and campaign outcomes. Estimated person-hours devoted to the campaign are 200–300 per year for campaign collaborators, 250–400 for extension agents and 150–250 for private sector personnel.

Calculation of reactive nitrogen losses. To obtain relevant nitrogen loss parameters, we conducted an exhaustive literature search of peer-reviewed publications using ISI-Web of Science (Thomson Reuters) and the China Knowledge Resource Integrated (CNKI) database. The literature search focused on field measurements of nitrogen losses, including NH_3 volatilization, NO_3^- leaching, N_2O emissions and nitrogen runoff in all major Chinese agricultural regions. All nitrogen losses had to have been measured both during field operations and throughout the entire growing season. The NH_3 volatilization had to have been measured within at least two weeks after nitrogen fertilization³⁰. The N_2O emissions had to have been measured daily using the static chamber technique for 7–10 days after nitrogen fertilization and for 3–10 days after other events that may have triggered N_2O gas emissions, such as rainfall, irrigation or tillage, as well as weekly or biweekly during the remaining periods³¹. Nitrogen leaching had to have been measured using the suction cup or lysimeter method³² or the soil sample method³³.

The final dataset consisted of 462 published references and 3,374 observations (see Supplementary Information for references). All analysed data were from the main agroecological zones. For data limitation, we combined northeast and northwest zones together as north China for reactive nitrogen loss calculations for the maize system, southeast and southwest zones together as south China for rice, and Yangtze River and southwest zones together as south China for wheat (Extended Data Figs 2–5). Across all zones and crops, the N_2O emissions, NO_3^- leaching, and nitrogen runoff (only paddy rice) increased exponentially with increasing nitrogen rate, whereas the relationship between NH_3 volatilization and nitrogen rate followed a linear model. The coefficient of nitrogen losses in response to nitrogen rate varied across the three zones and three crops, depending on climatic conditions, terrain, agricultural management practices (for example, with and without irrigation, cropping system) and soil types (Extended Data Figs 2–5, see Supplementary Information).

Using the exponential or linear models depicting the relationships between nitrogen losses and nitrogen rate (Extended Data Figs 2–5), we calculated N_2O emissions, NO_3^- leaching, nitrogen runoff and NH_3 volatilization relevant to ISSM interventions compared to control on the basis respective nitrogen rates. Total reactive nitrogen loss is reported as the sum of N_2O emissions, NO_3^- leaching, nitrogen runoff and NH_3 volatilization, expressed as kg reactive nitrogen per ha as well as yield-scaled reactive nitrogen loss (kg reactive nitrogen per Mg of grain produced).

Calculation of GHG emissions. The GHG emissions from the whole life cycle of crop production included CO_2 , CH_4 and N_2O ³⁴. The emissions consisted of three components⁶: those occurring after the application of nitrogen fertilizers, including direct and indirect N_2O emissions; those occurring during fertilizer manufacturing and transportation; and those from diesel fuel use in farming operations, such as sowing, tillage and harvesting. Indirect N_2O emissions after the application of nitrogen fertilizers were estimated through two indirect pathways via the volatilization of compounds, such as NH_3 and NO_x , with subsequent re-deposition downwind and N_2O emission there, and through leaching and runoff and subsequent N_2O emission downstream^{35,36}.

We also conducted an exhaustive literature search of peer-reviewed publications for relevant CH_4 emission parameters in the rice system using ISI-Web of Science and the CNKI database. The final dataset consisted of 85 published references and 464 observations according to the following criteria: the CH_4 emission data had to have been measured under field conditions, the measurement data had to have been conducted over an entire growth period of rice. For the rice system, the CO_2 equivalent of the CH_4 emission factor was 137 and 114 kg CH_4 ha⁻¹ for single-cropped rice in south and north China, respectively, and 212 kg CH_4 ha⁻¹ for double-cropped rice in south China. The 100-year global warming potential of CH_4 and N_2O were 25 and 298 times the intensity of CO_2 on a mass basis, respectively. In our work, soil CO_2 flux was not included because of data limitation and lower

impacts on a global scale³⁶. We calculated the climate footprint, expressed as kg CO_2 equivalent per ha, and as kg CO_2 equivalent per Mg grain.

Survey of prevailing farmer practices. To better understand what was happening in the Chinese agricultural landscape regarding productivity, resource management and various practices, a nationwide farmer survey was carried out during 2005–2014. A total of 1,944 counties were included, encompassing 66.4 million ha, which accounted for 73% of the total acreage planted for the three grains nationwide²⁷. In each county, 3–10 villages were selected; in each village, 30–120 farmers were randomly chosen as survey targets. The grand total of survey recipients, 8,630,079 individual farmers, included 2,891,694 farmers of maize, 3,505,004 farmers of rice and 2,233,381 farmers of wheat. The survey was conducted via face-to-face interviews by local (county and/or township) agricultural extension agents. The questionnaire was prescribed with non-open ended questions encompassing a variety of variables covering yield, crop varieties and fertilization practices (application rate, timing, product type, and so on). Only the yields and nitrogen rate data were extracted in the present study for the scenario analysis described below.

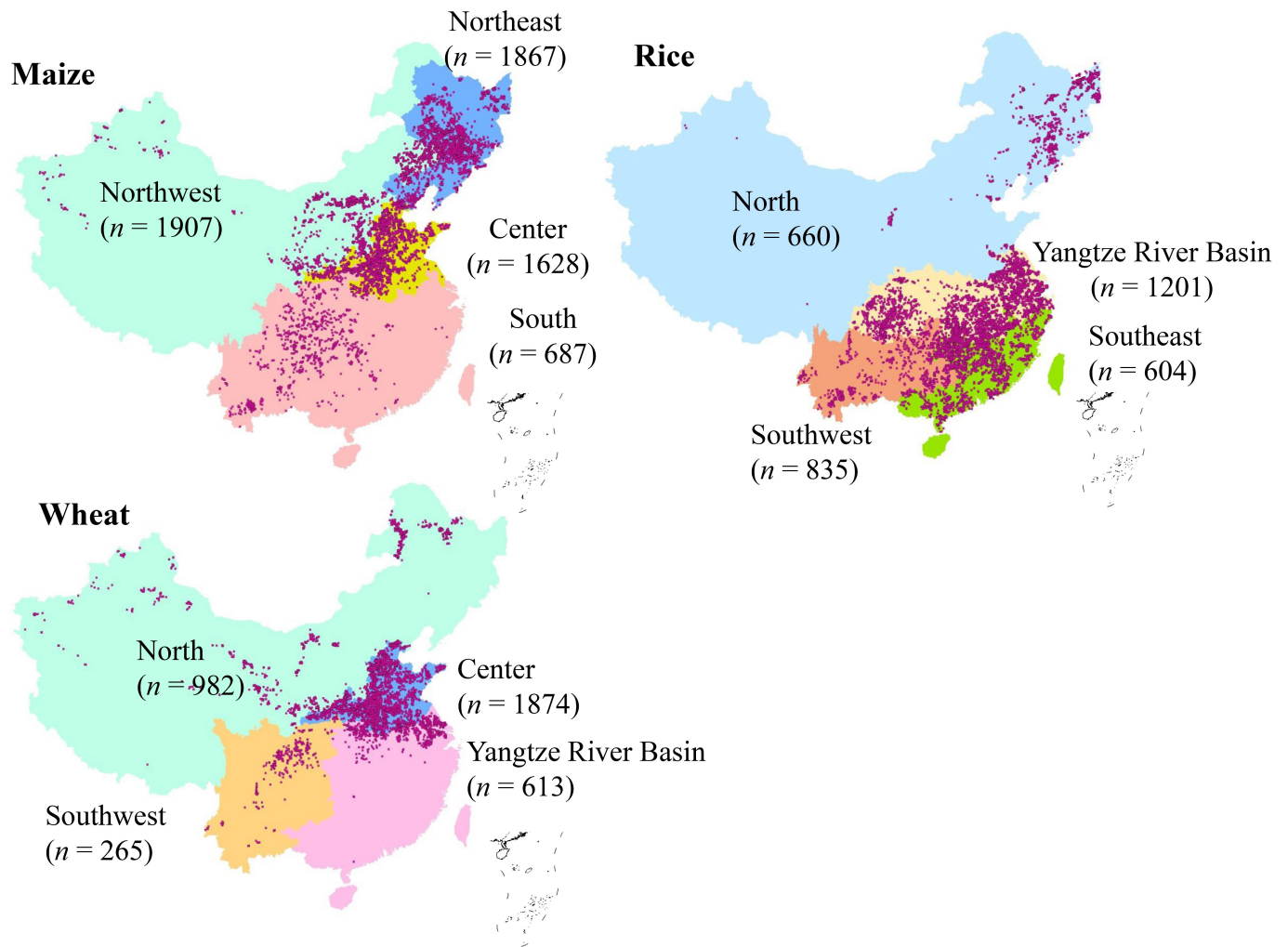
Scenario analysis. Considering the large variation in yield and nitrogen rate in the agroecological zones, we separately examined counties in their respective zones. For each crop zone, counties were grouped based on zone-average yield and nitrogen rate into high yield and high nitrogen, high yield and low nitrogen, low yield and low nitrogen, and low yield and high nitrogen (Extended Data Tables 2–4).

Scenario 1: counties with low yields and high nitrogen rates attain the benchmark (that is, ISSM-based yield and nitrogen rate). Scenario 2: counties with low yields and high nitrogen rates, and counties with low yields and low nitrogen rates achieve the benchmark. Scenario 3: counties with low yield and high nitrogen rates, low yields and low nitrogen rates, and high yields and high nitrogen rates attain the benchmark. The benchmarks were based on results from the field trials in the respective zones.

Data management. Raw data for the field trials (13,123 site years) and for the national campaign (37.7 million ha-years) were obtained and maintained by the network of campaign collaborators. At the same time, the raw data were reported to the head group at China Agricultural University, the campaign's lead institution, and entered into a database. The head group maintained the database and conducted summary analyses annually, which were provided to the funding agencies as well as campaign collaborators for feedback. For the current report of the 10-year span, all data analyses were performed at the China Agricultural University using the database. Data from the 13,123 site-year field trials were pooled; data analysis compared two treatments: ISSM-based intervention versus farmers' practices. Treatment effects were evaluated by one-way analysis of variance (ANOVA) using the Statistical Analysis System³⁷. Following *F*-tests in ANOVA, comparisons of means ($P < 0.05$) were made with a Fisher's protected least significant difference (LSD) test. For the national campaign, area-weighted means of crop yield, nitrogen rate, nitrogen productivity, reactive nitrogen losses, GHG emissions and the net economic gain were based on the data from the 14,600 paired sample pools.

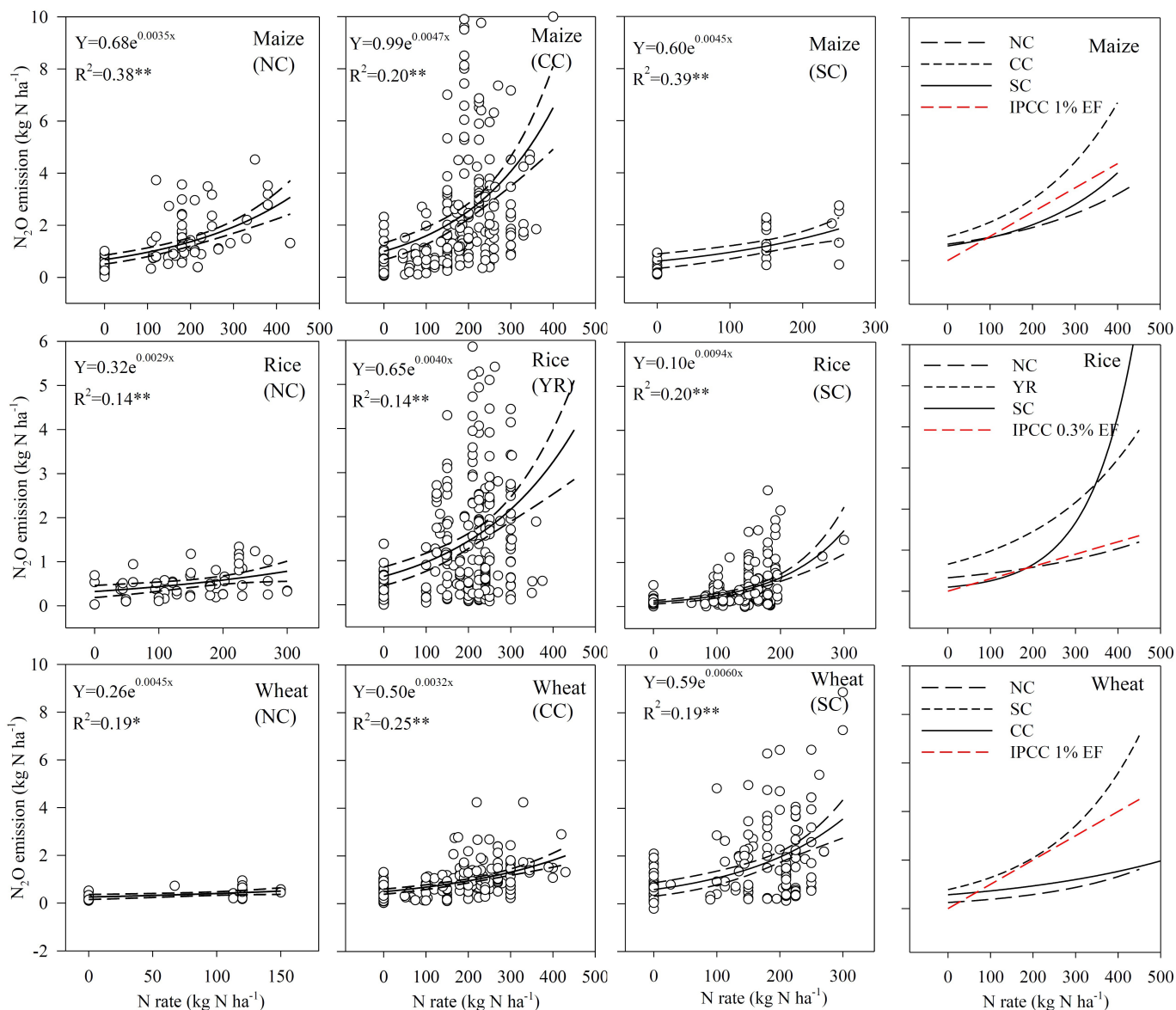
Data availability. All data are available from the corresponding authors upon reasonable request.

29. Li, Q., Huang, J. K., Luo, R. F. & Liu, C. F. China's labor transition and the future of China's rural wages and employment. *China World Econ.* **21**, 4–24 (2013).
30. Braschkat, J., Mannheim, T., Horlacher, D. & Marschner, H. Measurement of ammonia emissions after liquid manure application: I. Construction of a wind tunnel system for measurements under field conditions. *Z. Pflanz. Bodenkunde* **156**, 393–396 (1993).
31. Holland, E. et al. in *Standard Soil Methods for Long-Term Ecological Research* (eds Robertson, G. P. et al.), 185–201 (Oxford Univ. Press, 1999).
32. Lehmann, J. & Schroth, G. in *Trees, Crops and Soil Fertility — Concepts and Research Methods* (eds Schroth, G. & Sinclair, F. Ch. 7 151–165 (CAB International, 2003).
33. Zhao, R., Chen, X. & Zhang, F. Nitrogen cycling and balance in winter-wheat–summer-maize rotation system in Northern China. *Acta Pedol. Sin. (In Chinese)* **46**, 684–697 (2009).
34. Audsley, E. et al. Harmonisation of environmental life cycle assessment for agriculture. Final Report. Concerted Action MR3-CT94-2028. Silsoe Research Institute, Silsoe, UK (1997).
35. IPCC. *2006 IPCC Guidelines for National Greenhouse Gas Inventories, Prepared by the National Greenhouse Gas Inventories* (eds Eggleston, S. et al.) (Cambridge Univ. Press, 2006).
36. Smith, P. et al. in *Climate Change 2007: Mitigation of Climate Change* (eds Metz, B. et al.) (Cambridge Univ. Press, 2007).
37. SAS Institute. *SAS User's Guide: Statistics* (SAS Institute, 1998).



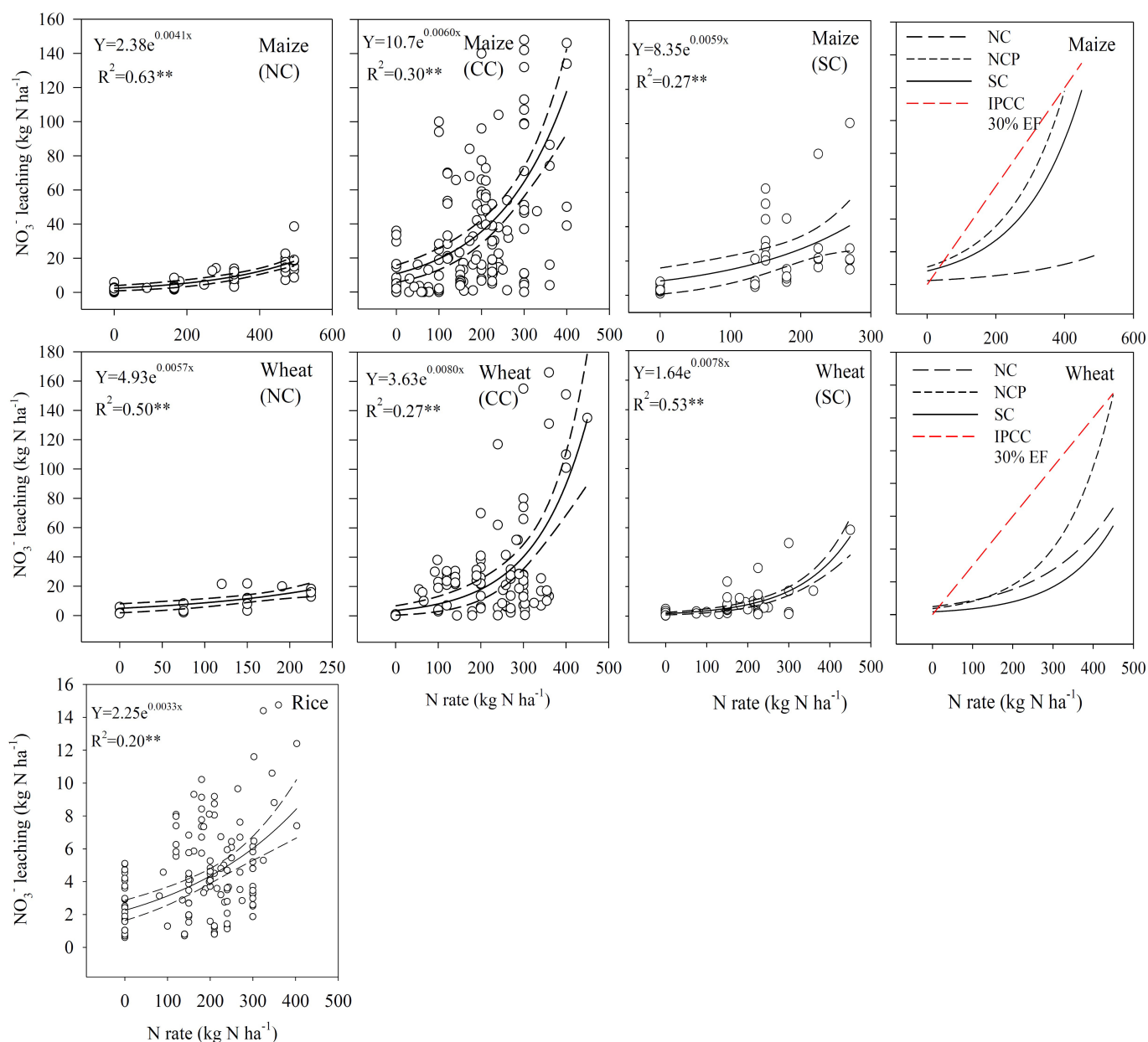
Extended Data Figure 1 | Distribution of field trials for maize, rice and wheat systems from 2005 to 2015 in China. The four coloured regions represent different agroecological zones for maize, rice and wheat.

Numbers in brackets are site years. Dots denote individual sites. The Chinese map was obtained from the Resource and Environment Data Cloud Platform (<http://www.resdc.cn/data.aspx?DATAID=202>).



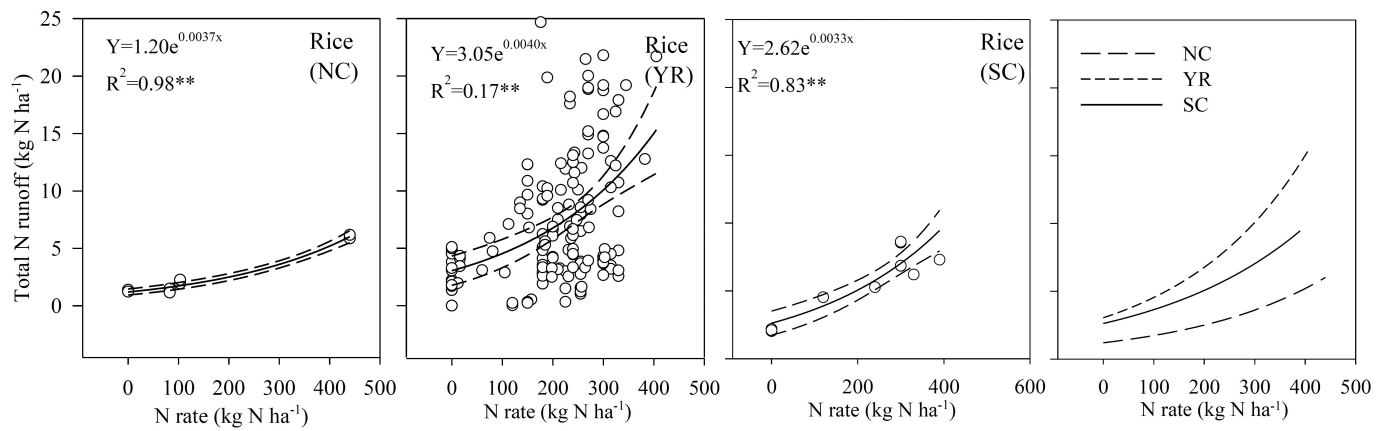
Extended Data Figure 2 | Exponential models describing the relationship between N_2O emissions and nitrogen rate. N_2O -N emissions were plotted against nitrogen rate for maize ($n = 417$), rice ($n = 740$) and wheat ($n = 395$). Red dotted lines are IPCC model-based calculations³⁵. NC, CC and SC refer to north China, central China and

south China, respectively, for maize and wheat production; NC-R, YR-R and SC-R refer to north China, Yangtze River Basin and south China, respectively, for rice production. $^{**}P < 0.01$ and $^*P < 0.05$ indicate the significance of the regression.

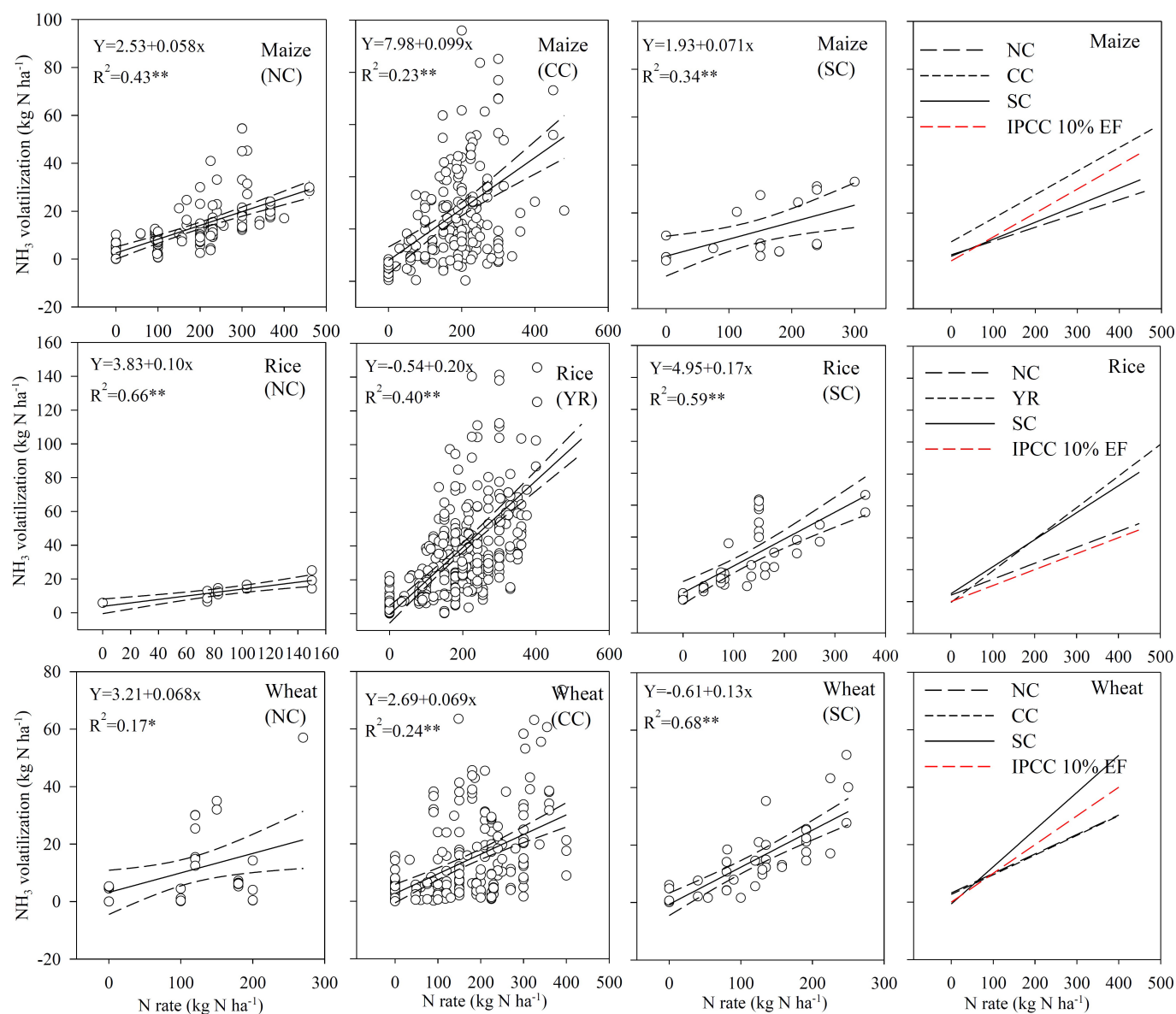


Extended Data Figure 3 | Exponential models describing the relationship between NO_3^- leaching and nitrogen rate. The NO_3^- leaching was plotted against nitrogen rate for maize ($n = 238$), rice ($n = 150$) and wheat ($n = 201$). The red dotted line is the IPCC model-

based calculation³⁵. NC, CC and SC refer to north China, central China and south China, respectively, for maize and wheat production. $^{**}P < 0.01$ indicates the significance of the regression.



Extended Data Figure 4 | Exponential models describing the relationship between nitrogen runoff and nitrogen rate for rice ($n = 216$). NC, YR and SC refer to north China, Yangtze River Basin and south China, respectively, for rice production.



Extended Data Figure 5 | Linear models describing the relationship between NH_3 volatilization and nitrogen rate. $\text{NH}_3\text{-N}$ volatilization was plotted against nitrogen rate for maize ($n=315$), rice ($n=423$) and wheat ($n=279$) growing seasons, respectively. The red dotted line is the IPCC

model³⁵. NC, CC and SC refer to north China, central China and south China, respectively, for maize and wheat production; NC, YR and SC refer to north China, Yangtze River Basin and south China, respectively, for rice production. $^{**}P < 0.01$.

Extended Data Table 1 | Conventional farmers' practice compared to ISSM-based recommendations for maize, rice or wheat in different agroecological zones

Region	Crops	Irrigation	Current practice	Critical points of ISSM-based recommendations
Northeast	Maize	Rainfed	Overuse N rate One-time fertilizer application Shallow tillage and poor planting Unsuitable variety Low density	Optimizing N with 180–200 kg N ha ⁻¹ Split before planting and around 6–8-leaf stage Deep tillage and improved planting High yielding variety with resistance to high density and disease Plant density in range of 6.5–7.0 plants m ⁻²
	Rice	Irrigation	Overuse N rate Too much in the early stage Poor planting quality with high seed Low density and more seeding per hole Unsuitable long-term flooding	Reducing N application rate by 20% Increasing the ratio of N application in the late growing season Reducing planting seed rate around 20% Increasing density by 20% and less seeding per hole by 30% Alternate wetting and moderate drying irrigation
Northwest	Maize	Rainfed	Low rainfall with water stress Unsuitable planting date and variety Low density	Increasing soil water retention capacity, and part of film mulching Optimizing sowing date and suitable variety Increased density with 4.8 to 8.5 plants m ⁻²
	Wheat	Rainfed	Low precipitation Low soil fertility Overuse N fertilizer Poor tillage and planting Unsuitable sowing rate and date	Increasing soil water retention capacity Straw return and applied organic manure Reduced N fertilizer application rate by 20% Deep tillage and improved planting Optimal sowing rate and date
Center	Maize	Irrigation	Unsuitable varieties Low density Overuse N and one-time use Harvest early	High-yielding varieties with resistance to high density and disease Increased density with 7.5 to 8.5 plants m ⁻² Optimal N rate and split in 6-leaf stage Harvest later by 5–7 days
	Wheat	Irrigation	Unsuitable sowing, early or late Shallow tillage and poor sowing quality Overuse N before planting Early side dressing in regreening stage Overuse and misused water management	Optimizing sowing date Improved sowing quality with deep tillage Optimal N application rate and 60% of total N use in shooting stage Optimal water management with rate, time, method
Yangtze River Basin	Rice	Irrigation	Overuse N and high use before planting Poor seeding quality Low density Unsuitable long-term flooding	Reduced N use before planting and increased N use in late season Improved seeding quality (reduce seed rate, control soil moisture, transplant seedlings in due time) Increased density Alternate wetting and moderate drying irrigation
	Wheat	Rainfed	Overuse N and high proportion applied in the early growing season Unsuitable sowing early or late Poor sowing quality with shallow tillage Hand broadcasting	Optimal N rate and high use in the mid-late season Suitable seeds sown at suitable time Improved sowing quality with deep tillage Mechanical sowing
South China	Rice	Rainfed	Low density by hand Overuse N and misuse PK fertilizer One-time fertilizer application Poor water and pest management	Increased density with 18–22 hole per m ² by machine Optimizing N and increased P and K use Split N fertilizer with 30–40% at sidedressing Improve water and pest management
	Maize	Rainfed	Low density with 3.7–4.5 plants m ⁻² Overuse N rate and high N losses Low soil fertility P and Zn deficiency	Increased density with 5–6 per m ² Optimal N rate with split N fertilization Increased soil fertility with organic manure Added P and Zn use before planting

Extended Data Table 2 | Maize yield, nitrogen rate, nitrogen productivity, reactive nitrogen losses and GHG emissions

Item	Planting area	Grain yield	N rate	N productivity	Nr losses	GHG emission
	Million ha	Mg ha ⁻¹	kg ha ⁻¹	kg kg ⁻¹	kg ha ⁻¹	kg CO ₂ eq ha ⁻¹
Northeast						
HH	3.16	9.63	217	45	22.4	3023
(n = 58)		(8.5-10.9)	(180-306)	(31-58)	(19-31)	(2611-4031)
HL	0.0352	9.15	147	66	16.6	2258
(n = 17)		(8.6-9.9)	(90-179)	(48-104)	(12-19)	(1639-2583)
LL	2.839	7.29	132	60	15.4	2088
(n = 51)		(3.7-8.3)	(45-179)	(23-183)	(8-19)	(1137-2653)
LH	1.355	8.03	206	40	21.5	2909
(n = 29)		(7.3-8.5)	(180-371)	(22-46)	(19-37)	(2595-5011)
Center China						
HH	2.419	7.93	257	31	88.9	4374
(n = 101)		(7.3-9.5)	(212-468)	(17-41)	(70-240)	(3577-9289)
HL	2.350	8.02	168 (101-	50	56.8	2964
(n = 110)		(7.2-11.4)	211)	(35-103)	(39-70)	(2033-3611)
LL	1.859	6.41	174	38	58.2	3038
(n = 102)		(3.0-7.2)	(70-212)	(24-80)	(33-70)	(1670-3662)
LH	1.907	6.47	256	26	87.9	4344
(n = 92)		(4.6-7.2)	(212-382)	(12-33)	(70-158)	(3580-7019)
Northwest						
HH	1.512	10.41	293	37	29.7	3967
(n = 102)		(8.3-15.2)	(218-510)	(22-59)	(22-55)	(3035-4850)
HL	0.805	9.83	170	61	18.4	2533
(n = 57)		(8.4-13.8)	(81-217)	(40-104)	(11-22)	(1604-3073)
LL	2.391	6.62	156	45	17.3	2369
(n = 136)		(3.0-8.2)	(57-217)	(25-113)	(10-22)	(1346-3057)
LH	0.893	6.97	273	26	27.6	3695
(n = 62)		(3.6-8.3)	(220-474)	(14-34)	(23-50)	(3048-6331)
South China						
HH	1.353	6.97	268	27	66	3886
(n = 121)		(6.7-10.7)	(208-582)	(11-48)	(47-310)	(3034-10273)
HL	0.704	6.87	171	41	38.6	2619
(n = 112)		(6.1-9.4)	(84-206)	(31-82)	(22-46)	(1607-3049)
LL	1.252	5.27	162	34	36.7	2484
(n = 164)		(3.2-6.1)	(79-207)	(19-67)	(22-46)	(1536-3024)
LH	0.991	5.44	252	22	59.9	3633
(n = 92)		(4.2-6.1)	(208-371)	(12-29)	(47-106)	(3033-5589)

Four categories are included: high yield and high nitrogen (HH), high yield and low nitrogen (HL), low yield and low nitrogen (LL), and low yield and high nitrogen (LH) (see Methods). Nr, reactive nitrogen. The values are means and ranges.

Extended Data Table 3 | Rice yield, nitrogen rate, nitrogen productivity, reactive nitrogen losses and GHG emissions

Item	Planting area	Grain yield	N rate	N productivity	Nr losses	GHG emission
	Million ha	Mg ha ⁻¹	kg ha ⁻¹	kg kg ⁻¹	kg ha ⁻¹	Kg CO ₂ eq ha ⁻¹
North						
HH	0.621	9.09	256	37	38.6	6557
(n = 52)		(8.3-10.7)	(188-525)	(17-48)	(30-79)	(5885-9487)
HL	0.541	8.97	143	66	24.3	5440
(n = 27)		(8.3-10.5)	(74-185)	(47-122)	(12-19)	(4717-5841)
LL	1.361	7.55	130	63	22.8	5322
(n = 63)		(6.0-8.3)	(35-185)	(42-171)	(12-29)	(4400-5836)
LH	0.143	7.45	238	33	36.2	6368
(n = 23)		(5.5-8.2)	(189-406)	(17-42)	(30-59)	(5910-7993)
Yangtze River						
HH	2.768	8.36	270	32	71.2	8183
(n = 110)		(7.5-9.7)	(189-426)	(19-47)	(50-115)	(7151-10523)
HL	2.566	8.03	154	53	41.9	6730
(n = 106)		(7.5-9.1)	(98-188)	(41-79)	(29-50)	(6063-7168)
LL	4.330	6.78	148	47	40.6	6666
(n = 182)		(2.9-7.5)	(50-188)	(16-142)	(18-50)	(5550-7136)
LH	1.119	6.92	227	32	60.4	7649
(n = 48)		(5.7-7.5)	(189-512)	(14-39)	(50-144)	(7148-11922)
Southeast						
HH	1.483	7.21	203	36	49.9	9102
(n = 672)		(6.6-10.7)	(169-367)	(20-52)	(43-87)	(8658-11830)
HL	2.089	7.07	146	50	38.154	8411
(n = 76)		(6.6-7.9)	(60-168)	(40-111)	(21-43)	(7462-8689)
LL	2.029	6.03	147	42	38.2	8419
(n = 78)		(4.9-6.6)	(84-168)	(30-76)	(26-43)	(7725-8682)
LH	1.021	6.07	189	32	46.8	8916
(n = 52)		(4.9-6.6)	(169-236)	(21-38)	(43-57)	(8646-9476)
Southwest						
HH	0.257	8.34	225	38	54.4	9362
(n = 48)		(7.3-10.5)	(180-304)	(27-46)	(45-72)	(8699-10525)
HL	0.521	8.07	144	58	37.7	8373
(n = 46)		(7.3-9.8)	(69-176)	(45-120)	(23-44)	(7599-8754)
LL	0.717	6.42	144	46	37.6	8364
(n = 64)		(4.8-7.3)	(63-176)	(32-114)	(22-44)	(7483-8764)
LH	0.424	6.45	222	30	53.7	9319
(n = 40)		(4.6-7.3)	(180-348)	(17-39)	(45-82)	(8728-11392)

Four categories are included: high yield and high nitrogen, high yield and low nitrogen, low yield and low nitrogen, and low yield and high nitrogen (see Methods). The values are means and ranges.

Extended Data Table 4 | Wheat yield, nitrogen rate, nitrogen productivity, reactive nitrogen losses and GHG emissions

Item	Planting area	Grain yield	N rate	N productivity	Nr losses	GHG emission
	Million ha	Mg ha ⁻¹	kg ha ⁻¹	kg kg ⁻¹	kg ha ⁻¹	Kg CO ₂ eq ha ⁻¹
North						
HH	1.000	6.17	264	25	47.1	3275
(n = 103)		(4.7-12.1)	(172-534)	(11-45)	(29-146)	(2235-6826)
HL	0.443	5.68	132	46	23.3	1850
(n = 39)		(4.7-7.6)	(65-171)	(29-93)	(15-28)	(1194-2228)
LL	1.568	3.11	99	40	19.3	1508
(n = 122)		(1.0-4.6)	(14-169)	(12-122)	(10-28)	(664-2211)
LH	0.205	4.06	214	19	35.6	2696
(n = 19)		(2.8-4.5)	(172-279)	(13-26)	(29-47)	(2232-3599)
Yangtze River						
HH	1.540	5.73	219	27	39.9	3499
(n = 66)		(4.7-7.2)	(171-405)	(16-41)	(30-97)	(2735-7370)
HL	0.541	5.74	137	45	23.4	2304
(n = 31)		(4.8-9.3)	(50-166)	(29-121)	(9-28)	(1319-2735)
LL	0.615	3.39	128	28	21.9	2168
(n = 71)		(1.2-4.6)	(45-169)	(15-58)	(8-29)	(1167-2687)
LH	0.145	3.88	223	18	40.8	3567
(n = 18)		(1.9-4.6)	(173-346)	(5-25)	(30-73)	(2745-5778)
Center China						
HH	3.565	7.09	269	27	55.6	4195
(n = 119)		(6.5-8.6)	(222-465)	(20-52)	(40-187)	(3630-6769)
HL	3.538	7.01	187	39	33.2	3288
(n = 87)		(6.5-10.9)	(91-222)	(30-120)	(17-40)	(2202-3679)
LL	2.517	5.44	166	34	29.3	3063
(n = 94)		(2.4-6.4)	(56-221)	(18-81)	(13-40)	(1887-3662)
LH	1.518	5.95	268	23	56.0	4172
(n = 59)		(3.5-6.4)	(223-495)	(13-28)	(41-229)	(3679-7341)
Southwest						
HH	0.376	4.58	174	27	30.3	2788
(n = 56)		(3.5-7.6)	(139-251)	(16-55)	(24-46)	(2303-3957)
HL	0.439	4.43	117	40	19.9	2020
(n = 45)		(3.6-6.6)	(35-138)	(27-120)	(7-23)	(1070-2321)
LL	0.526	2.71	971	33	16.6	1766
(n = 80)		(1.1-3.5)	(19-138)	(11-130)	(4-23)	(852-2238)
LH	0.054	2.78	194	15	36.3	3174
(n = 41)		(1.4-3.5)	(140-516)	(6-25)	(24-171)	(2256-11502)

Four categories are included: high yield and high nitrogen, high yield and low nitrogen, low yield and low nitrogen, and low yield and high nitrogen (see Methods). The values are means and ranges.

Extended Data Table 5 | Area-weighted yield, nitrogen rate, total amounts of grain output, nitrogen fertilizer use, reactive nitrogen losses, and GHG emissions with scenario analysis using a 3-step progression, compared to prevailing practices, that is, business as usual

Item	unit	BAU	S1	S2	S3
Yield	Mg ha ⁻¹	6.98	7.28 (104%)	8.05 (115%)	8.23 (118%)
N rate	kg N ha ⁻¹	195	186 (96%)	197 (102%)	177 (91%)
Crop production	Mt	464	483 (104%)	534 (115%)	546 (118)
N use	Mt	12.9	12.4 (96%)	13.1 (102%)	11.8 (91%)
Nr losses	Mt	2.79	2.64 (94%)	2.70 (99%)	2.35 (84%)
GHG emission	Mt	307	299 (97%)	304 (99%)	284 (92%)

Business as usual (BAU) practices were calculated using county averages from the farmers' surveys and planting acreage from national statistical data. Scenario 1 (S1): counties in the low yield and high nitrogen category (see Methods, Extended Data Tables 2–4) attaining ISSM-based yield and nitrogen rate. Scenario 2 (S2): counties in low yield and low nitrogen category, in addition to those in S1; Scenario 3 (S3): counties in high yield and high nitrogen category plus those in S2. Values in parentheses indicate relevant scenario outcomes as a percentage of BAU.

Pervasive phosphorus limitation of tree species but not communities in tropical forests

Benjamin L. Turner¹, Tania Brenes-Arguedas¹ & Richard Condit¹

Phosphorus availability is widely assumed to limit primary productivity in tropical forests^{1,2}, but support for this paradigm is equivocal³. Although biogeochemical theory predicts that phosphorus limitation should be prevalent on old, strongly weathered soils^{4,5}, experimental manipulations have failed to detect a consistent response to phosphorus addition in species-rich lowland tropical forests^{6–9}. Here we show, by quantifying the growth of 541 tropical tree species across a steep natural phosphorus gradient in Panama, that phosphorus limitation is widespread at the level of individual species and strengthens markedly below a threshold of two parts per million exchangeable soil phosphate. However, this pervasive species-specific phosphorus limitation does not translate into a community-wide response, because some species grow rapidly on infertile soils despite extremely low phosphorus availability. These results redefine our understanding of nutrient limitation in diverse plant communities and have important implications for attempts to predict the response of tropical forests to environmental change.

One of the longest-standing paradigms in ecology is that productivity in tropical forests is limited by phosphorus (P) availability^{1,2}. The paradigm is supported by biogeochemical theory, which states that P depletion during long-term pedogenesis is sufficient to limit productivity on the old, strongly weathered soils that characterize much of the tropical biome^{4,5}. There is also a wealth of indirect evidence for P limitation in tropical forests, including high nitrogen (N) availability¹⁰, high N-to-P ratios in leaves¹ and correlations between forest properties and soil fertility at continental scale^{11,12}. However, evidence from nutrient-addition experiments in tropical forests is scarce and contradictory. A community-wide growth response to P has been observed in a monodominant forest in Hawaii¹³, but not in species-rich lowland tropical forests in Africa, Southeast Asia and the neotropics^{6–9}, and a recent meta-analysis found that the overall evidence for P limitation in the tropics is largely inconclusive³.

Here we combine data on tree growth rates, species distributions and soil phosphatase activities to precisely quantify P limitation of individual species and whole communities in lowland tropical forests. We define P limitation as faster growth at greater P availability, which can manifest at the level of an individual species or an entire community. We measured the growth of 18,970 individual trees that were ≥ 10 mm in diameter at breast height (dbh; the trunk diameter at 1.3 m above the ground surface), comprising 541 species occurring in a network of 32 forest-dynamics plots across the Isthmus of Panama (Supplementary Table 1). The plots vary in size from 1 to 50 ha and were censused at least twice to provide growth rates for individual stems. The plot network spans a rainfall gradient (1,870–3,280 mm) with marked variation in lithology and soils, which generates a steep natural gradient in P availability that is unrelated to rainfall^{14,15}. In particular, readily exchangeable phosphate extracted by anion-exchange membranes (resin phosphate)—a sensitive measure of the power of the soil to supply P for biological uptake—varies more than 300-fold¹⁴, which represents a similar range to phosphate availability in lowland tropical forests globally^{16–18}.

We modelled species-specific growth rates using hierarchical models to disentangle the influence of environmental variables from the confounding effect of species turnover across the gradients of P and rainfall¹⁴. This approach isolates the influence of individual variables (that is, fixed effects) on the growth of the average species, given a hypothetical scenario in which other variables are held constant and the average species exists in all locations. Growth rates increased significantly with increasing resin phosphate (likelihood ratio test (LRT) for the fixed effect of resin phosphate, $P < 0.0001$; Fig. 1a, b, Extended Data Fig. 1 and Extended Data Table 1). Responses were independent of tree size (resin phosphate \times dbh interaction, LRT $P = 0.72$), which indicates that both large and small trees grew faster in response to higher concentrations of resin phosphate (Fig. 1a, b). At intermediate soil moisture, the predicted growth of an average 100-mm-dbh tree increased from 0.77 mm y^{-1} at the lowest resin phosphate concentration to 1.03 mm y^{-1} at the highest, a growth increase of 34%. The growth of an average 10-mm-dbh tree across the same P gradient increased from 0.15 to 0.18 mm y^{-1} , an increase of 20%. The model indicated significant variation among species in their response to P (random effects for resin phosphate, LRT $P = 0.0015$; Extended Data Table 2), as demonstrated by the negative responses of some species to increasing concentrations of resin phosphate (Fig. 1a, b). However, the significant fixed effect of resin phosphate demonstrates that most species respond positively to increasing P availability. Indeed, 90% of common species (that is, with >20 individuals in the dataset) responded positively to P as large trees, and 84% responded positively as small trees; only a small number of species did not respond positively to P in either life history stage (Extended Data Fig. 2a).

By contrast, increasing soil moisture increased growth rates only for smaller trees, consistent with smaller trees suffering greater water stress than adults owing to a less extensive root system (fixed effect of moisture, LRT $P = 0.003$; inclusion of a dbh \times moisture interaction parameter, LRT $P = 0.002$; Fig. 1c, d and Extended Data Table 1). Therefore, across the range of soil moisture deficit in our study area, the predicted growth of a 10-mm-dbh tree at intermediate soil P increased from 0.15 mm y^{-1} at the driest site to 0.24 mm y^{-1} at the wettest site (LRT for a model using only trees of 10–100 mm dbh, $P = 0.001$). The growth of a 100-mm-dbh tree did not change significantly across the moisture gradient (0.86 – 0.89 mm y^{-1} ; LRT for a model using only trees of ≥ 100 mm dbh, $P = 0.1$), reflecting the erratic responses of individual species to moisture (Fig. 1c, d).

Resin phosphate in our plots was not correlated with total inorganic N or soil properties, such as organic matter (total carbon (C) and total N) or texture (for example, clay concentration), but was correlated positively with base cations (Supplementary Table 2). To investigate the influence of other nutrients on tree growth, we performed separate model runs using N, calcium (Ca), potassium (K) and the micro-nutrient manganese (Mn) in place of resin phosphate (Extended Data Table 2). Neither Mn nor total inorganic N or K, the two most important plant nutrients other than P, were significant predictors of tree growth rates. Calcium was significant when it was the only nutrient

¹Smithsonian Tropical Research Institute, Apartado 0843-03092, Balboa, Ancon, Panama.

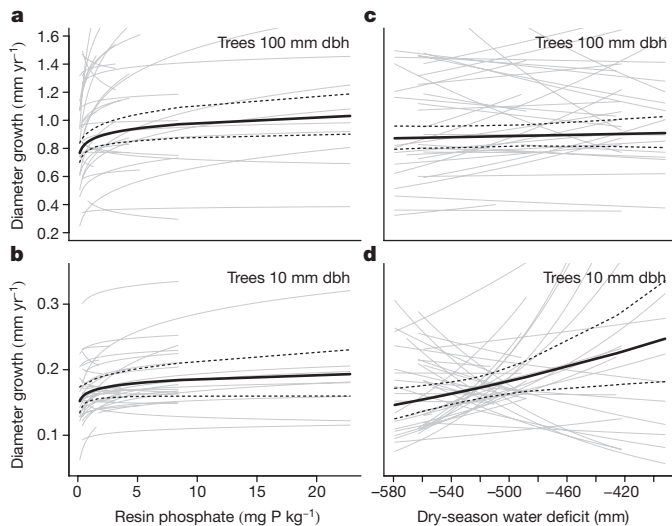


Figure 1 | Tree growth responses to phosphorus and moisture.

a–d, Growth responses of the average species to resin phosphate (**a**, **b**) and dry-season soil-moisture deficit (**c**, **d**), predicted by the hierarchical model for an average large tree of 100 mm dbh (**a**, **c**) and an average small tree of 10 mm dbh (**b**, **d**). Phosphorus responses are predicted at average dry-season soil-moisture deficit, and moisture responses are predicted at average resin phosphate concentration. Dashed black lines show 95% credible intervals, calculated as the 2.5th and 97.5th quantiles of predictions from 1,000 random draws of model parameters. Grey lines represent the predicted responses of abundant species where they occur along the phosphorus gradient. For large trees (**a**, **c**), these are the 40 most-abundant species with dbh ≥ 100 mm. For small trees (**b**, **d**), these are the 40 most-abundant species with dbh between 10 and 50 mm. Some fast-growing species exceed the upper boundary of the y axis. $n = 18,970$ individual trees and 541 species were used in the models.

in the model, but was not significant in a model that included resin phosphate (Extended Data Table 3). Although we cannot rule out the possibility that Ca has an independent influence on growth rates in our plots, our model results therefore indicate that P is the primary nutrient determining tree growth rates in the lowland tropical forests of Panama. Indeed, there is little evidence that Ca limits productivity in forested ecosystems, including tropical forests¹, although we recognize that Ca limitation is possible in some tropical regions—including parts of Amazonia and Southeast Asia—that have soils with concentrations of exchangeable base cations at least an order of magnitude lower than in most of our plots^{16,17}.

Model predictions and piecewise linear regression demonstrate that growth responses to P increase markedly below approximately 2 mg P kg⁻¹ resin phosphate (Fig. 1a, b and Extended Data Fig. 2b). Strong P limitation below this threshold is supported by changes in the activity of soil phosphatase enzymes, which release phosphate from organic compounds and are synthesized by plants and microbes in response to P demand¹⁹. For 83 soils across the P gradient (Supplementary Table 3), including the 32 plots analysed for tree growth, the activity of phosphomonoesterase and phosphodiesterase—the two enzymes involved in the hydrolysis of the majority of the organic P in tropical forest soils¹⁵—decreased exponentially with increasing resin phosphate (Fig. 2a and Extended Data Fig. 2c). Phosphatase activity increased markedly below 2 mg P kg⁻¹ resin phosphate, which is almost identical to the concentration that triggers phosphatase genes and other phosphate starvation responses in bacteria²⁰ (0.16 μ M P, equivalent to 2.15 mg P kg⁻¹ resin phosphate). This supports previously published evidence that P availability can constrain the activity of soil microbes in lowland tropical forests^{21,22} and, together with tree growth responses, demonstrates a coherent threshold for strong P limitation above- and belowground.

The resin phosphate threshold can be quantified precisely by changes in tree community composition along the P gradient. The distributions

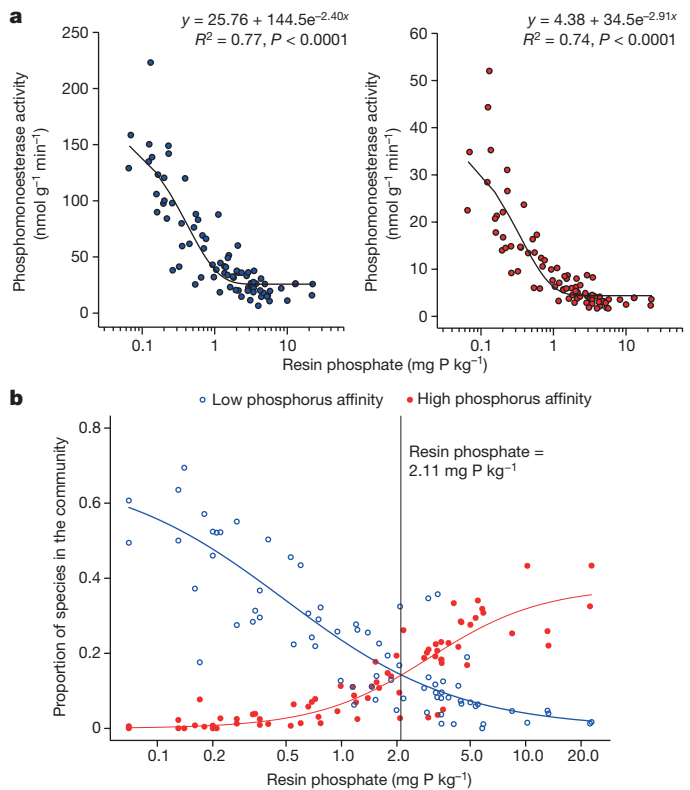


Figure 2 | A threshold for strong phosphorus limitation in lowland tropical forests.

a, The relationship between resin phosphate (logarithmic scale) and phosphatase activities in soils from 83 sites under lowland tropical forest in Panama, showing marked increases in phosphomonoesterase activity (left, blue circles) and phosphodiesterase activity (right, red circles) at low concentrations of resin phosphate. The hydrolysis product is methylumbelliferone and the model fits are negative exponential functions determined by nonlinear regression. **b**, The relationship between resin phosphate and the proportion of tree species associated with low levels of soil phosphorus (effect sizes < -0.8 , open blue circles and blue line) or high levels of soil phosphorus (effect sizes > 0.8 , closed red circles and red line) in 72 lowland tropical forests in Panama. The proportion of species with low phosphorus affinity equals the proportion of species with high phosphorus affinity at 2.11 mg P kg⁻¹ resin phosphate. The models are sigmoidal fits and were derived by nonlinear regression.

of individual tree species in lowland forests of Panama are determined primarily by P and moisture availability¹⁴. The strength of the association between a species and a resource can be described quantitatively by its effect size, the first-order parameter of the logistic model describing the relationship between the occurrence of a species and the resource¹⁴. A positive effect size for P indicates that a species occurs predominantly on high-P soils, whereas a negative effect size indicates that a species occurs predominantly on low-P soils. The current study included 364 species with significant low-P associations (effect size < -0.8 ; 66% of the entire community) and 58 species with significant high-P associations (effect size > 0.8 ; 11% of the entire community) (Supplementary Table 4). Sites with low resin phosphate are dominated by species with low-P affinity, but these species are gradually replaced along the P gradient by species with high-P affinity. The resin phosphate concentration at which the tree community contains equal proportions of species with low- and high-P-affinity is 2.11 mg P kg⁻¹ (Fig. 2b). This concentration is similar only for widespread species (Extended Data Fig. 2d), and varying the definition of significant P affinity (using effect sizes between ± 0.5 and ± 1.0) and the minimum number of occurrences of a species (between three and eight plots) yields resin phosphate threshold values between 1.96 and 2.37 mg P kg⁻¹.

Despite evidence for a coherent signature of strengthening P limitation below 2 mg P kg⁻¹ resin phosphate, the growth rates for individual

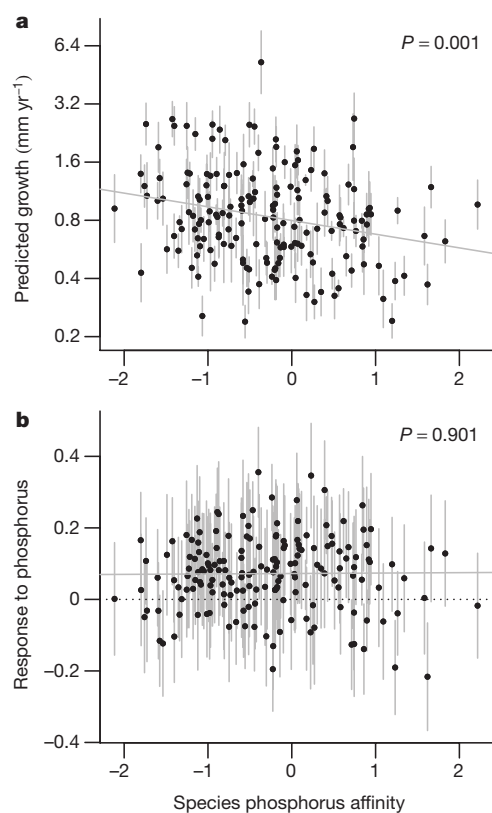


Figure 3 | Predicted growth rates and growth responses to increasing soil phosphorus for individual common species as a function of their phosphorus affinity. **a, b,** Growth rates (**a**) and phosphorus responses (**b**) were predicted by the hierarchical model. Common species are defined as those with growth data for >20 individuals. Each point represents the response ± standard error of a single species as estimated from the model, assuming a tree of 100 mm dbh. Growth rates in **a** are estimated at intermediate moisture and resin phosphate levels, although relative values were similar when estimated at low or high levels of phosphorus (Extended Data Fig. 3a). Responses in **b** were estimated at average soil moisture, and positive y-axis values indicate that the species grows faster at higher resin phosphate concentration. The *P* values indicate the significance of linear regressions of random effect size against species phosphorus affinity (grey solid line).

species estimated by the hierarchical model at intermediate moisture and resin phosphate levels were on average greater for species associated with low-P soils (Fig. 3a). This pattern is consistent for modelled growth rates estimated at low or high resin phosphate concentrations, and for observed growth rates across the natural species ranges (Extended Data Fig. 3a, b). However, species P affinities were not related to their growth responses to P (Fig. 3b), demonstrating that individual species respond to P in a similar manner irrespective of where they occur on the P gradient.

Even though most individual species are P limited and grow faster as P availability increases, community-wide growth rates (at the plot level) did not vary significantly across the P gradient (Fig. 4). Similarly, neither plot-level aboveground biomass nor relative biomass increment varied significantly with resin phosphate (Extended Data Fig. 3c). It therefore appears that overall growth is maintained on infertile soils by a subset of species that grow rapidly despite extremely low P availability (Fig. 3a). These species gradually disappear from the community as soil P increases (Fig. 1a, b), replaced by species that are better adapted to higher P availability but that have slightly slower growth on average. The net result is consistent community-wide growth across the entire P gradient despite widespread species-specific P limitation.

These results redefine our concept of nutrient limitation in species-rich plant communities by demonstrating that P limitation

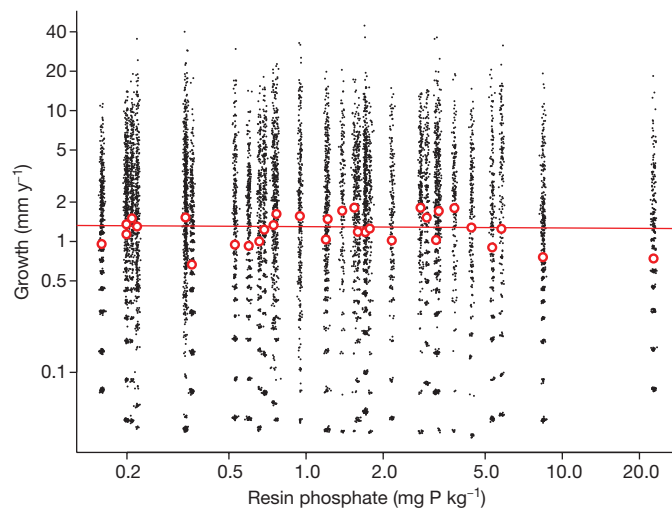


Figure 4 | Observed community-wide growth rates as a function of resin phosphate concentration. Plot-level growth rates (red circles) are for trees ≥ 100 mm dbh in 32 plots in lowland tropical forests in Panama. Growth rates of individual trees are shown as black points. Both axes are log-transformed. The line shows a standard linear regression between log-transformed growth and log-transformed resin phosphate; the slope is slightly negative (-0.022), but not significantly different from zero (*P* = 0.06).

occurs at the level of the individual species rather than the entire community. Where diversity is high, variation in fertility drives species turnover and differences in community composition, rather than variation in growth rates of the same species assemblage. This pattern is common along fertility gradients in species-rich plant communities^{23,24} and supports the suggestion that retrogression—the process by which low P availability causes a decline in biomass and productivity on old soils⁵—is unlikely to occur in diverse tropical forests because they contain species that can be productive on low-P soils²⁵.

Although we cannot explain how some species maintain high growth rates on low-P soils, it presumably involves mechanisms that promote efficient use of P, including exhaustive re-translocation of foliar P, synthesis of sulfolipids or galactolipids instead of phospholipids, low ribosomal RNA concentrations or efficient exploitation of soil organic P compounds^{1,26,27}. These low-P-specialist species are therefore potential targets for efforts to develop crops that can maintain growth on infertile soils. Although some species might be particularly sensitive to high P availability^{23,26}, as suggested by the negative response to P of some species in our plots, most low-P specialists respond positively to small increases in resin phosphate within their natural ranges. Their exclusion from high-P sites is therefore presumably driven by physiological or ecological factors such as responses to herbivory or pathogens²⁸, or trade-offs between growth and nutrient acquisition²⁹, which paradoxically cause them to be outcompeted by slower-growing species that are better-adapted to survive and reproduce on more fertile soils.

Our results have implications for efforts to incorporate P into coupled climate–carbon cycle models to improve predictions for the tropical biome under future atmospheric chemistry and climate scenarios³⁰. For example, P constraints on growth responses to increasing atmospheric carbon dioxide concentrations are likely to be species specific, confounding the simple inclusion of P limitation in earth system models. However, in addition to revealing the nature of P limitation in tropical forests, we show a quantitative resin phosphate threshold below which P limitation strengthens markedly above- and belowground. Although resin phosphate data are not widely available for lowland tropical forest soils, the values correspond closely to ‘plant-available’ P concentrations measured by routine soil P tests and are correlated strongly with total and organic P concentrations

(Extended Data Figs 4, 5), making it possible to predict the pan-tropical extent of P limitation in lowland tropical forests. Given that extractable P concentrations below 2 mg P kg^{-1} occur widely in Asia, Africa and South America^{16–18}, it seems likely that species-specific P limitation is pervasive in tropical forests worldwide.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 December 2016; accepted 26 January 2018.

Published online 7 March 2018.

- Vitousek, P. M. Litterfall, nutrient cycling, and nutrient limitation in tropical forests. *Ecology* **65**, 285–298 (1984).
- Vitousek, P. M. & Sanford, R. L. Jr. Nutrient cycling in moist tropical forest. *Annu. Rev. Ecol. Syst.* **17**, 137–167 (1986).
- Cleveland, C. C. *et al.* Relationships among net primary productivity, nutrients and climate in tropical rain forest: a pan-tropical analysis. *Ecol. Lett.* **14**, 939–947 (2011).
- Walker, T. W. & Syers, J. K. The fate of phosphorus during pedogenesis. *Geoderma* **15**, 1–19 (1976).
- Wardle, D. A., Walker, L. R. & Bardgett, R. D. Ecosystem properties and forest decline in contrasting long-term chronosequences. *Science* **305**, 509–513 (2004).
- Mirmanto, E., Proctor, J., Green, J., Nagy, L. & Suriantata. Effects of nitrogen and phosphorus fertilization in a lowland evergreen rainforest. *Philos. Trans. R. Soc. Lond. B* **354**, 1825–1829 (1999).
- Alvarez-Clare, S., Mack, M. C. & Brooks, M. A direct test of nitrogen and phosphorus limitation to net primary productivity in a lowland tropical wet forest. *Ecology* **94**, 1540–1551 (2013).
- Wright, S. J. *et al.* Potassium, phosphorus, or nitrogen limit root allocation, tree growth, or litter production in a lowland tropical forest. *Ecology* **92**, 1616–1625 (2011).
- Newbery, D. M. *et al.* Does low phosphorus supply limit seedling establishment and tree growth in groves of ectomycorrhizal trees in a Central African rainforest? *New Phytol.* **156**, 297–311 (2002).
- Brookshire, E. N. J., Gerber, S., Menge, D. N. L. & Hedin, L. O. Large losses of inorganic nitrogen from tropical rainforests suggest a lack of nitrogen limitation. *Ecol. Lett.* **15**, 9–16 (2012).
- Quesada, C. A. *et al.* Basin-wide variations in Amazon forest structure and function are mediated by both soils and climate. *Biogeosciences* **9**, 2203–2246 (2012).
- ter Steege, H. *et al.* Continental-scale patterns of canopy tree composition and function across Amazonia. *Nature* **443**, 444–447 (2006).
- Vitousek, P. M. & Farrington, H. Nutrient limitation and soil development: experimental test of a biogeochemical theory. *Biogeochemistry* **37**, 63–75 (1997).
- Condit, R., Engelbrecht, B. M. J., Pino, D., Pérez, R. & Turner, B. L. Species distributions in response to individual soil nutrients and seasonal drought across a community of tropical trees. *Proc. Natl Acad. Sci. USA* **110**, 5064–5068 (2013).
- Turner, B. L. & Engelbrecht, B. M. J. Soil organic phosphorus in lowland tropical rain forests. *Biogeochemistry* **103**, 297–315 (2011).
- Quesada, C. A. *et al.* Variations in chemical and physical properties of Amazon forest soils in relation to their genesis. *Biogeosciences* **7**, 1515–1541 (2010).
- Palmiotto, P. A. *et al.* Soil-related habitat specialization in dipterocarp rain forest tree species in Borneo. *J. Ecol.* **92**, 609–623 (2004).
- Yang, X. & Post, W. M. Phosphorus transformations as a function of pedogenesis: a synthesis of soil phosphorus data using Hedley fractionation method. *Biogeosciences* **8**, 2907–2916 (2011).
- Marklein, A. R. & Houlton, B. Z. Nitrogen inputs accelerate phosphorus cycling rates across a wide variety of terrestrial ecosystems. *New Phytol.* **193**, 696–704 (2012).
- Torriani-Gorini, A. in *Phosphate in Microorganisms: Cellular and Molecular Biology* (eds Torriani-Gorini, A. *et al.*) 1–4 (ASM, 1994).
- Cleveland, C. C. & Townsend, A. R. Nutrient additions to a tropical rain forest drive substantial soil carbon dioxide losses to the atmosphere. *Proc. Natl Acad. Sci. USA* **103**, 10316–10321 (2006).
- Turner, B. L. & Wright, S. J. The response of microbial biomass and hydrolytic enzyme activities to a decade of nitrogen, phosphorus, and potassium addition in a lowland tropical rain forest. *Biogeochemistry* **117**, 115–130 (2014).
- Wassen, M. J., Venterink, H. O., Lapshina, E. D. & Tanneberger, F. Endangered plants persist under phosphorus limitation. *Nature* **437**, 547–550 (2005).
- Zemunik, G., Turner, B. L., Lambers, H. & Laliberté, E. Diversity of plant nutrient-acquisition strategies increases during long-term ecosystem development. *Nat. Plants* **1**, 15050 (2015).
- Kitayama, K. Comment on 'Ecosystem properties and forest decline in contrasting long-term chronosequences'. *Science* **308**, 633 (2005).
- Lambers, H. *et al.* Proteaceae from severely phosphorus-impooverished soils extensively replace phospholipids with galactolipids and sulfolipids during leaf development to achieve a high photosynthetic phosphorus-use-efficiency. *New Phytol.* **196**, 1098–1108 (2012).
- Turner, B. L. Resource partitioning for soil phosphorus: a hypothesis. *J. Ecol.* **96**, 698–702 (2008).
- Bagchi, R. *et al.* Pathogens and insect herbivores drive rainforest plant diversity and composition. *Nature* **506**, 85–88 (2014).
- Zalamea, P.-C. *et al.* Seedling growth responses to phosphorus reflect adult distribution patterns of tropical trees. *New Phytol.* **212**, 400–408 (2016).
- Reed, S. C., Yang, X. & Thornton, P. E. Incorporating phosphorus cycling into global modeling efforts: a worthwhile, tractable endeavor. *New Phytol.* **208**, 324–329 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank R. Pérez, S. Aguilar and the many field assistants who helped in collection of the plot data; staff at the STRI Soils Laboratory for assistance in the collection and analysis of soils; and J. Dalling, E. Laliberté, M. Sheldrake and G. Zemunik for comments on the manuscript.

Author Contributions R.C. collected tree growth data, B.L.T. collected soil data, T.B.-A. and R.C. conducted statistical analysis, and B.L.T. wrote the manuscript with input from T.B.-A. and R.C.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to B.L.T. (TurnerBL@si.edu).

METHODS

Tree growth rates. We measured the growth of 18,970 stems comprising 541 species (excluding palms and lianas) in a network of 32 plots across the Isthmus of Panama^{14,31–33}. Annual precipitation across the plots varied between 1,870 and 3,280 mm y^{-1} , and elevation ranged between 20 and 643 m above sea level, although most plots were below 200 m above sea level. Details of the floristic composition of the plots were previously published^{14,31,32} and locations are detailed in Supplementary Table 1. In each plot, all trees larger than 100 mm dbh were tagged, measured and identified to species. Twenty-six of the plots also included stems between 10 and 100 mm dbh, typically inside a central 40 × 40 m quadrat. Lianas were not included in the census. Most plots were 1 ha in area, but two plots were larger (6 ha at Sherman and 50 ha on Barro Colorado Island, BCI). To prevent these two larger plots from dominating the results with their much larger sample size, we subsampled two 1-ha plots from each larger plot with all trees to 100 mm dbh, and a 40 × 40 m quadrat within each plot to 10 mm dbh.

Tree growth was measured between 1996 and 2011. Twenty-five plots were established between 1996 and 1999, six plots were established after 1999 and one plot (BCI) is older (the first census was completed in 1983). Most plots were re-censused only once after their establishment. We included only census intervals greater than three years and ignored intermediate re-censuses (Supplementary Table 1). For consistency, for the BCI plot we included only data from between 2000 and 2010. For five of the plots that had multiple census intervals greater than three years, we calculated the growth in each census interval independently and averaged growth over the whole period to yield a single measure for each individual stem. All tree census data, including every tree measurement in every census through to the middle of 2012, is available online without restriction in permanent archives at the Smithsonian Institution Library^{34,35}.

We calculated growth as diameter increment per year (mm y^{-1}) for the main stem of each individual tree, as long as the stem survived and the dbh was taken at an identical position in both censuses. Extreme errors were excluded on the basis of an independent assessment of measurement error: trees were eliminated if dbh in the second census was more than four times the measurement error lower than dbh in the first census, and when growth was >75 mm y^{-1} . We also eliminated palms, for which diameter growth has little meaning for most species. Growth data was log-transformed for analysis. Modest negative growth rates (that is, those not excluded as extreme errors) were included by converting all growth rates ≤ 0 to half the minimum measurable growth (0.5 mm divided by the census time interval). Because census intervals were large for some of the plots, we calculated the mean dbh of each individual tree over the whole census interval. For the analysis, we log-transformed dbh and centred it at 100 mm. We included a quadratic term for dbh in the model.

Soil analysis. Soils in the plots include several taxonomic orders (oxisols, ultisols, alfisols and inceptisols)¹⁵ and vary considerably in chemical properties, including pH (3.3–7.0), organic C (2–10%) and resin phosphate (0.16–22.8 mg P kg^{-1}) (Supplementary Table 1). A number of soil parameters have been measured in the plots and previously reported in detail^{14,15,33}. For each site, soil data represent the average of five cores (inventory transects and 40 × 40 m plots), 13 cores (1-ha forest-dynamics plots) or 25 cores (the 6-ha plot at Sherman and the 50-ha plot on BCI), where each core was analysed individually. Cores were taken from the surface soil (0–10 cm in depth), which integrates the nutrient cycle and contains the majority of the extractable nutrients and fine roots. Additional samples were taken up to a depth of 1 m in the soil profile, but extractable nutrient concentrations, especially of resin phosphate, were much lower at depth and are not discussed further. The 32 plots studied in the hierarchical analysis are a subset of the broader plot network, and a number of additional plots were studied for phosphatase activities and to calculate P and moisture effect sizes (see later; Supplementary Table 3).

Our primary measure of available P was readily-exchangeable phosphate determined by extraction with anion-exchange membranes (that is, resin phosphate)³⁶. This is a sensitive measure of the capacity of the soil to supply phosphate that reflects the distribution of approximately 60% of the species in the region¹⁴. The detection limit of the method is approximately two orders of magnitude lower than other procedures for estimating plant-available phosphate, such as Mehlich-III³⁷ or Bray-1³⁸, and extraction is conducted in deionized water, which avoids soil-specific chemical interactions that occur with acidic or chelating extractants. For the hierarchical analysis, resin phosphate concentrations were log-transformed and standardized to a mean of zero.

Extractable base cations (including Ca, K and magnesium (Mg)), and micro-nutrients (including iron (Fe), Mn and zinc (Zn)) were determined by Mehlich-III extraction³⁷ and inductively-coupled plasma optical-emission spectrometry on an Optima 7300DV (Perkin Elmer). Phosphate was determined in the Mehlich extracts by automated molybdate colorimetry. Extractable phosphate was also determined by Bray-1 extraction and molybdate colorimetry³⁸. Inorganic

N (including ammonium and nitrate) and dissolved organic N (DON) were determined by extraction in 2 M KCl with colorimetric detection³⁹. Total P was determined by ignition (550 °C × 1 h) and extraction (1 M H₂SO₄, 1:50 soil-to-solution ratio, 16 h extraction), with phosphate detection by automated molybdate colorimetry^{15,40}. This procedure provides a simple and rapid estimate of total P for most soils, although it can underestimate the true values in strongly weathered soils such as many of those in the current study. On average, resin phosphate represented only a small fraction of the total P ($0.5 \pm 0.05\%$), although with considerable variation (range 0.04–1.95%). Organic P was determined by alkaline extraction¹⁵ in a solution containing 0.25 M NaOH and 0.05 M EDTA. Phosphatase activities were determined using methylumbelliferyl-linked substrates as previously described⁴¹. Assays were conducted at pH 5.0 in 50 mM acetate buffer for phosphomonoesterase and phosphodiesterase, the two enzymes involved in the hydrolysis of the majority of the organic P in tropical forest soils¹⁵. Activity was expressed on the basis of dry soil, microbial biomass C and total organic C (determined as total C; no soil contained carbonate). Microbial C was determined by fumigation–extraction⁴² and total C by dry combustion on a Flash 1112 elemental analyser (Thermo Fisher Scientific).

All analyses, other than total C and total P, were determined on fresh soils within 4 h (KCl extraction) or 24 h (resin phosphate and phosphatase activity) of sampling, to avoid the rapid changes that can occur during storage or pretreatment³⁶. All soil chemical properties are expressed on the basis of oven-dry equivalent soil (determined by drying at 105 °C for 24 h).

Dry-season water deficit. To characterize site moisture status we calculated dry-season water deficit (in mm) as previously described^{14,33}. Dry-season water deficit measures the intensity of the dry season (between December and April) as the net moisture deficit: cumulative daily precipitation minus evapotranspiration at its most extreme at the end of the dry season. A more negative water deficit indicates a stronger (that is, longer) dry season. For the analysis, water deficit was standardized and centred on a median of -525 mm; positive values represent more humid sites and negative values represent drier sites.

Data analysis. All data manipulation and analysis was done using R statistical software (R Development Core Team; <https://www.r-project.org/>). We investigated whether dry-season soil water deficit and resin phosphate levels influenced forest growth at the community and at the species level using hierarchical linear mixed-effect models ('lmer' function in the package 'lme4')⁴³. The hierarchical modelling approach is a multiple regression analysis that differentiates between fixed effects (the overall response to a parameter) and random effects (the random variation in responses among individuals). In our models we evaluated the fixed effects of moisture, tree size and nutrients, and random variation in responses among species and among plots.

Model selection and evaluation of parameter probability values was performed using LRTs and the Akaike information criterion (AIC). LRTs evaluate whether the likelihood of a model fit changes significantly as each individual parameter is added (or dropped) from the model. If the LRT is significant, and the AIC is smaller when a parameter is added to the model, that parameter is considered a significant improvement to the model. Therefore, probability values represent the probability that the parameter improves model fit relative to the same model without that parameter. A significant fixed effect in the hierarchical model indicates an overall response to a parameter (for example, resin phosphate), whereas a significant random effect indicates that the response to the fixed effect varies across species or plots. The variance and standard deviation values for the random effects terms indicate the extent to which species vary in response to the fixed effect, and are not an estimate of the variation explained by the term.

Data entered the model as the log-transformed growth of each individual tree. In the final model, each species was allowed to vary randomly in its response to all model parameters, and random parameters were allowed to co-vary with one another. All analysis and figures were conducted based on the following form (in R notation): $\log \text{ growth} \sim \log \text{ dbh} + \log \text{ dbh squared} + \text{moisture deficit} + \log \text{ resin P} + \text{moisture deficit}:\log \text{ dbh} + (\log \text{ dbh} + \log \text{ dbh squared} + \text{moisture deficit} + \log \text{ resin P} + \text{moisture deficit}:\log \text{ dbh} \mid \text{species})$.

We evaluated the model using the complete tree community data, including 541 species and 32 plots. All species-specific model outputs, including estimated growth and responses to moisture and P, are shown in Supplementary Table 5. However, as in many tropical forests, most species were rare and appeared as singletons or at a very low sample size. These rare species contribute to the overall model, but the species-specific estimates have reduced significance because the hierarchical model 'shrinks' them towards the community mean. Therefore, to look at species-specific trends we evaluated 175 species that had growth data for at least 20 individuals in the study sites (Supplementary Table 6).

Diameter growth is highly correlated with basal area growth, but we checked whether basal area growth responded differently than dbh growth to P in simplified models that included P, moisture deficit and dbh as predictors of growth. In these

models, the fixed effect of P and the responses of individual species using the two growth estimates were essentially identical ($R^2 = 0.86$), and in fact showed a slightly stronger effect of P on basal area growth than on dbh growth.

We also tested a model with plot as a random effect, to account for factors such as stem density or canopy structure that were not included in the model, but which could contribute to differences in growth among plots. Adding the plot random effect significantly improved the model, but the overall results were essentially unchanged, confirming that for most species growth increases with increasing P even after accounting for random unexplained variation among plots (Extended Data Table 4a). Compared to the main model, the growth response to P was slightly stronger in the model with a plot random effect, and the fixed effect of P was significantly improved ($P = 0.02$). However, the model output was noisier, as expected from the inclusion of an additional plot-level parameter, with the lower t -value suggesting greater variability in the responses among species (Extended Data Table 4a). We therefore did not include a random plot-effect in the final model to avoid over-parameterization.

Stem density in our plots is greater at wetter sites (Supplementary Table 1), as is common elsewhere in the tropics^{11,44}, and plot-wide growth rates are negatively correlated with the density of trees ≥ 10 cm dbh owing to greater shade in sites with many large stems. After controlling for moisture, stem density is not correlated significantly with resin phosphate or any other soil nutrient. However, to confirm whether stem density affects the relationship between growth and P, we ran a separate model that included stem density as a fixed effect (Extended Data Table 4b). Stem density slightly improved the original model, but the effect of P on growth remained highly significant (LRT $P < 0.0001$). Furthermore, stem density did not improve the model containing a plot random effect (described earlier), presumably because the plot effect includes variation in stem density as well as other unidentified differences among plots. Variation in stem density therefore does not affect our conclusions regarding the relationship between P and growth.

Variation in growth within species was considerably larger than variation among species or due to environmental factors, so the model explained only about 35% of the variance in growth. Of this, the largest effect was the change in growth rate due to tree size. For example, at average dry-season soil water deficit and resin phosphate concentration, a tree of 10 mm dbh grew 0.17 mm y^{-1} , compared to 0.87 and 1.25 mm y^{-1} for trees of 100 and 200 mm dbh, respectively. Nonetheless, we were able to detect significant community-wide trends in tree growth as a function of the environment.

Although P and rainfall are not correlated in our plot network, P is correlated with base cations. We therefore conducted separate linear mixed effects models evaluating the effect of different parameters on tree growth, evaluated as $\log(\text{growth})$. These models included resin phosphate, total inorganic N (ammonium plus nitrate) and the base cations Ca and K. The linear models were run using the function `lmer` (from the `lme4` R package) and probability (P) values were calculated using LogLik model comparisons after dropping or adding one parameter at a time (Extended Data Tables 2, 3).

Classification of P and moisture affinities. The majority of species in our plots have restricted distribution across the study area, with many showing affinity towards the wetter or the drier side of the gradient, or for high or low concentrations of resin phosphate. Moisture and P affinities were quantified by logistic regression, involving the assessment of species-by-species occurrence probability with respect to dry-season moisture deficit (moisture affinity) and resin phosphate (P affinity)¹⁴. The presence or absence of each species in 72 locations across the Isthmus of Panama was fitted in a hierarchical model as a function of dry-season water deficit and soil parameters. The strength of the affinity of each species to moisture or P is defined by the 'effect size', which is the first-order parameter of the logistic model. Species with strong negative P associations (that is, negative effect sizes) occur predominantly in sites with low resin phosphate concentrations, whereas species with strong positive P associations (that is, positive effect sizes) occur predominantly in sites with high resin phosphate concentrations. Species moisture and P affinities (effect sizes) are listed in Supplementary Table 4.

Because P affinity is a quantitative response, dividing species into groups is arbitrary. We classed low-P specialists as species with logistic coefficients (effect sizes) < -0.8 and high-P specialists as species with coefficients > 0.8 . Species with P association scores between -0.8 and 0.8 occur across the entire range of resin phosphate concentrations and were classified as generalists. Relaxing or strengthening the definition of significant effect size to values between ± 0.5 and ± 1.0 did not influence the principal results. Similarly, moisture affinity is the first-order slope of the effect of dry-season water deficit in this function and represents the probability of occurrence in wetter or drier sites. Effect sizes greater than zero represent species that are found mostly in wetter sites (that is, sites with shorter dry seasons) and values less than zero represent species that are more frequently found in drier sites (that is, sites with longer dry seasons).

To assess the influence of resin phosphate on species distributions, we summed the total number of species at every inventory site that had significant responses to soil P. Both specialist groups essentially disappear at one end of the P gradient: at low resin phosphate concentrations there are no high-P-specialist species, whereas at high resin phosphate concentrations there are no low-P-specialist species. Fitting sigmoidal models to the data yielded a value at which the two models intersect—the point at which the tree community consists of equal proportions of species with high-P affinities and species with low-P affinities (Fig. 2b).

Biomass and biomass growth. Aboveground dry biomass (AGB) was estimated for each of the plots using allometric equations relating volume to stem diameter, combined with species-specific wood density. Details and examination of errors were previously published⁴⁵. A caveat is that several early censuses omitted measurements of very large buttressed trunks owing to the difficulty of transporting long ladders to remote sites. Omitting large trees can cause a substantial error in the calculation of standing biomass, but we used later censuses to avoid this bias. In the end, only two large trees were omitted, from two different plots, accounting for no more than 1.5% of total forest mass. Biomass growth posed a greater concern, because this requires two censuses in which every tree is measured at precisely the same position on the stem. To avoid this bias, we measured relative biomass growth on the sample of trees measured twice at the same position (that is, total biomass increment of all those surviving trees divided by initial biomass of those same trees). This enabled all plots to be included. Growth and standing stock of biomass were regressed against the estimated intensity of the dry season and the logarithm of resin phosphate across the 32 plots.

Some of the forest surrounding the Panama Canal is successional, and has been re-growing since the United States took over the region in the early part of the 20th century. However, we do not have precise information on successional status of individual plots other than those on BCI (over 500 years since disturbance). Based on the presence of individuals of gap-demanding or edge species, four of the thirty-two plots support forest that appears to be relatively young (less than 60 years of regrowth), and all the rest are at least 120 years old. However, the four plots were included in biomass calculations because they did not have lower biomass than mature secondary or primary forest on BCI. This is consistent with growth rates and aboveground biomass of secondary forests in the region approaching those of undisturbed forest after a few decades of regrowth⁴⁶.

Piecewise linear regression. We used piecewise linear regression to investigate whether the forest as a whole showed a nonlinear growth response to soil P (that is, that growth rates increase (or decrease) faster at one end of the P gradient than the other). To confirm that the decreasing slope indicated by model predictions is real (that is, not an artefact of log-transformation), we employed a piecewise regression model, using the response variable $\log(\text{growth})$ and the predictor resin phosphate (untransformed). Piecewise regression fits a standard linear response in two separate sections of the x axis and determines whether the slope of y (the response variable) differs between the sections. The key notion is that the break point, b_x , defining the two sections, $x < b_x$ and $x > b_x$, is estimated along with the two separate slopes. The null hypothesis is that the response of y to x is linear across the entire range of x , with no change in slope. That hypothesis is rejected if there is any b_x that allows the slopes on either side to differ significantly. The piecewise model includes the constraint that separate linear regressions in the two sections meet at the break point, so the two regressions are not fully independent. However, there is no direct constraint on the position of the break or the two slopes. If the underlying response y on x is not linear, the piecewise method should demonstrate the manner in which it differs.

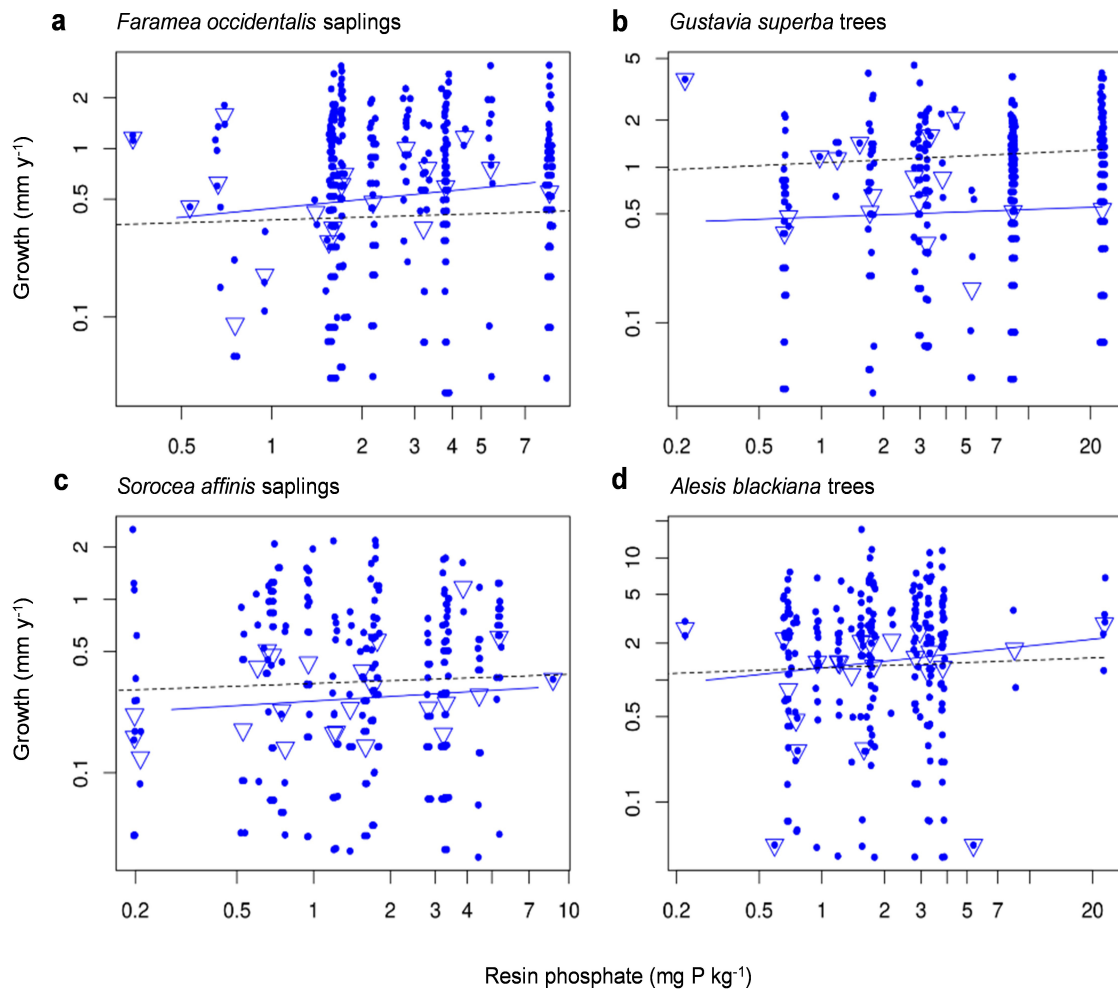
Application of the piecewise model applied to the growth of many tree species in response to P involved a substantial number of species that spanned only narrow ranges of P, especially those restricted to low P. Because we were interested in whether there was a universal P concentration at which the response of growth changes, we were forced to pose the question about only generalist (occurring across a reasonably wide range of P levels) species. We thus restricted the analysis to those species occurring over at least a tenfold variation in P. Because two different regressions are needed for each species, the model is data-demanding, so we added the further restriction that species have at least 20 individuals at five or more sites. Furthermore, the piecewise model employed only P (and not moisture) as the predictor of growth.

The piecewise model with two sections requires four parameters: b_x , the two slopes s_1 and s_2 , and a single intercept, y_0 , defined as the estimated response at b_x . As in all the models we employed, species were incorporated as a random effect. A full model thus included these four parameters for every species, plus a set of hyper-parameters (the fixed effect) describing the mean response across species. A likelihood function describing the probability of observations given all the parameters, assuming Gaussian error functions, is required. Parameters were fitted using a previously described Bayesian hierarchical method¹⁴. We were interested

in the fixed effect, b_x , s_1 , and s_2 , and whether the slopes differ significantly, as established if 95% intervals of posterior distributions did not overlap.

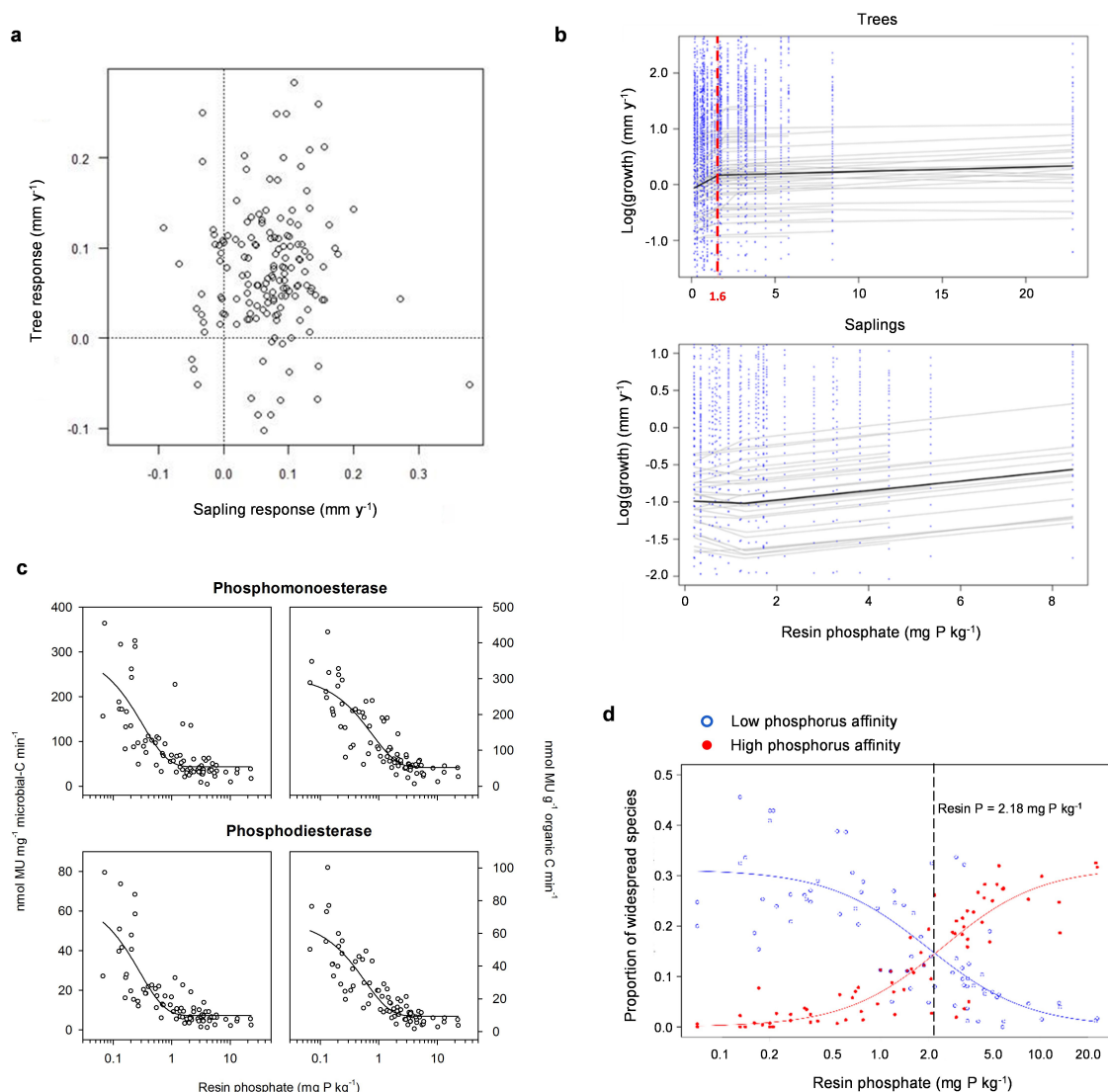
Data availability. Tree census data are available online from the Smithsonian Library at <http://dx.doi.org/10.5479/data.bci.20130603> and <http://dx.doi.org/10.5479/data.stri.2016.0622>. Site and soils data and species responses generated during hierarchical modelling are available in the Supplementary Information. All other data are available from the corresponding authors upon reasonable request.

31. Pyke, C. R., Condit, R., Aguilar, S. & Lao, S. Floristic composition across a climatic gradient in a neotropical lowland forest. *J. Veg. Sci.* **12**, 553–566 (2001).
32. Condit, R. *et al.* Tropical forest dynamics across a rainfall gradient and the impact of an El Niño dry season. *J. Trop. Ecol.* **20**, 51–72 (2004).
33. Engelbrecht, B. M. J. *et al.* Drought sensitivity shapes species distribution patterns in tropical forests. *Nature* **447**, 80–82 (2007).
34. Condit, R. *et al.* Barro Colorado forest census plot data, 2012 version. Center for Tropical Forest Science Databases <http://dx.doi.org/10.5479/data.bci.20130603> (2013).
35. Condit, R., Pérez, R., Aguilar, S. & Lao, S. Data from tree censuses and inventories in Panama. Center for Tropical Forest Science Databases <http://dx.doi.org/10.5479/data.stri.2016.0622> (2016).
36. Turner, B. L. & Romero, T. E. Short-term changes in extractable inorganic nutrients during storage of tropical rain forest soils. *Soil Sci. Soc. Am. J.* **73**, 1972–1979 (2009).
37. Mehlich, A. Mehlich 3 soil test extractant: a modification of Mehlich 2 extractant. *Commun. Soil Sci. Plant Anal.* **15**, 1409–1416 (1984).
38. Bray, R. H. & Kurtz, L. T. Determination of total, organic and available forms of phosphorus in soils. *Soil Sci.* **59**, 39–45 (1945).
39. Mulvaney, R. L. in *Methods of Soil Analysis, Part 3: Chemical Methods* (eds Sparks D. L. *et al.*) 1123–1184 (Soil Science Society of America, 1996).
40. Walker, T. W. & Adams, A. F. R. Studies on soil organic matter: I. influence of phosphorus content of parent materials on accumulations of carbon, nitrogen, sulfur, and organic phosphorus in grassland soils. *Soil Sci.* **85**, 307–318 (1958).
41. Turner, B. L. Variation in pH optima of hydrolytic enzyme activities in tropical rain forest soils. *Appl. Environ. Microbiol.* **76**, 6485–6493 (2010).
42. Vance, E. D., Brookes, P. C. & Jenkinson, D. S. An extraction method for measuring soil microbial biomass C. *Soil Biol. Biochem.* **19**, 703–707 (1987).
43. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
44. Anderson-Teixeira, K. J. *et al.* CTFS-ForestGEO: a worldwide network monitoring forests in an era of global change. *Glob. Chang. Biol.* **21**, 528–549 (2015).
45. Chave, J. *et al.* Error propagation and scaling for tropical forest biomass estimates. *Philos. Trans. R. Soc. Lond. B* **359**, 409–420 (2004).
46. Batterman, S. A. *et al.* Key role of symbiotic dinitrogen fixation in tropical forest secondary succession. *Nature* **502**, 224–227 (2013).



Extended Data Figure 1 | Growth responses of individual species to resin phosphate. **a–d**, Blue points represent the observed growth of individual trees, and blue triangles the species mean growth in a plot. The solid blue line is the modelled species response to resin phosphate, and the dashed black line is the fixed response of the entire community (as in Fig. 1). The four species are among the most abundant and widespread in the two size classes: *Faramaea occidentalis* (Rubiaceae), an understory evergreen tree/shrub (**a**); *Sorocea affinis* (Moraceae), an understory deciduous tree (**b**); *Gustavia superba* (Lecythidaceae), an understory

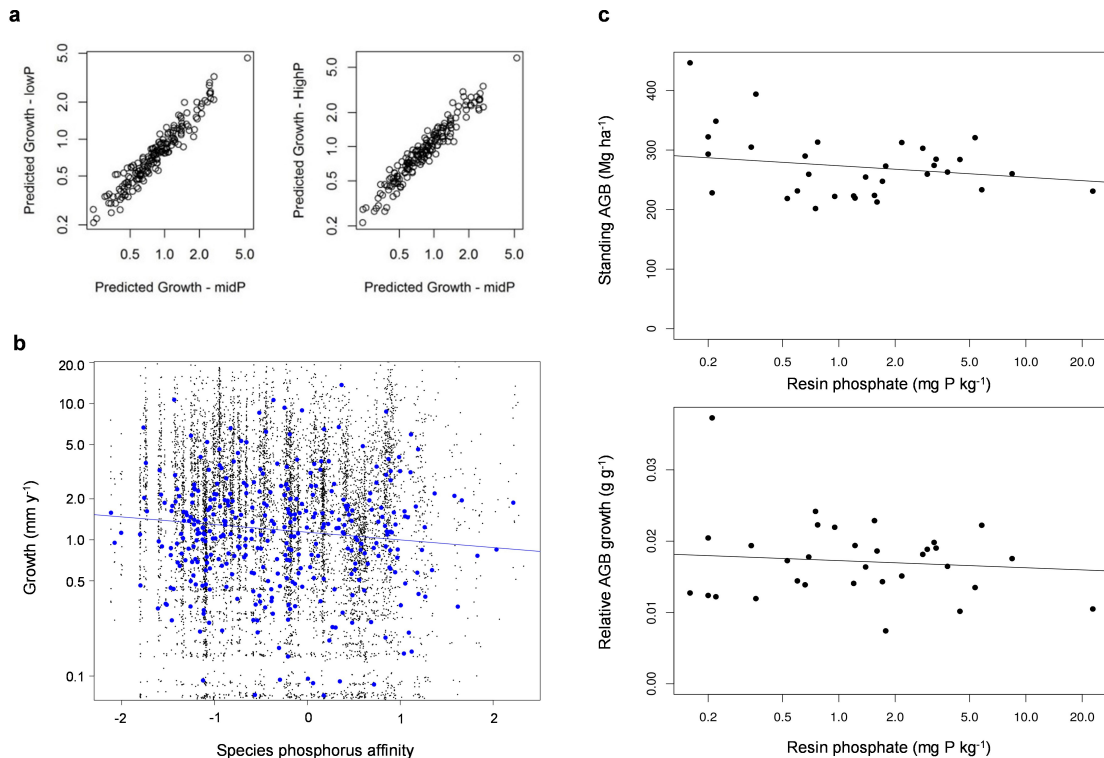
tree (**c**); and *Alseis blackiana* (Rubiaceae), a canopy tree (**d**). Saplings (**a**, **c**) include all individuals ≥ 10 mm and < 50 mm dbh; the dashed black line is the community-wide estimate at 30 mm dbh. Trees (**b**, **d**) include all individuals ≥ 100 mm dbh; the dashed black line is the community-wide estimate at the mean dbh of all trees ≥ 100 mm dbh. Both the y axes (growth in mm y^{-1}) and x axes (resin phosphate in mg P kg^{-1}) are plotted on logarithmic scales. The number of individuals were: 398 saplings (*F. occidentalis*), 328 saplings (*S. affinis*), 620 trees (*G. superba*) and 253 trees (*A. blackiana*).



Extended Data Figure 2 | Responses to resin phosphate above and below ground.

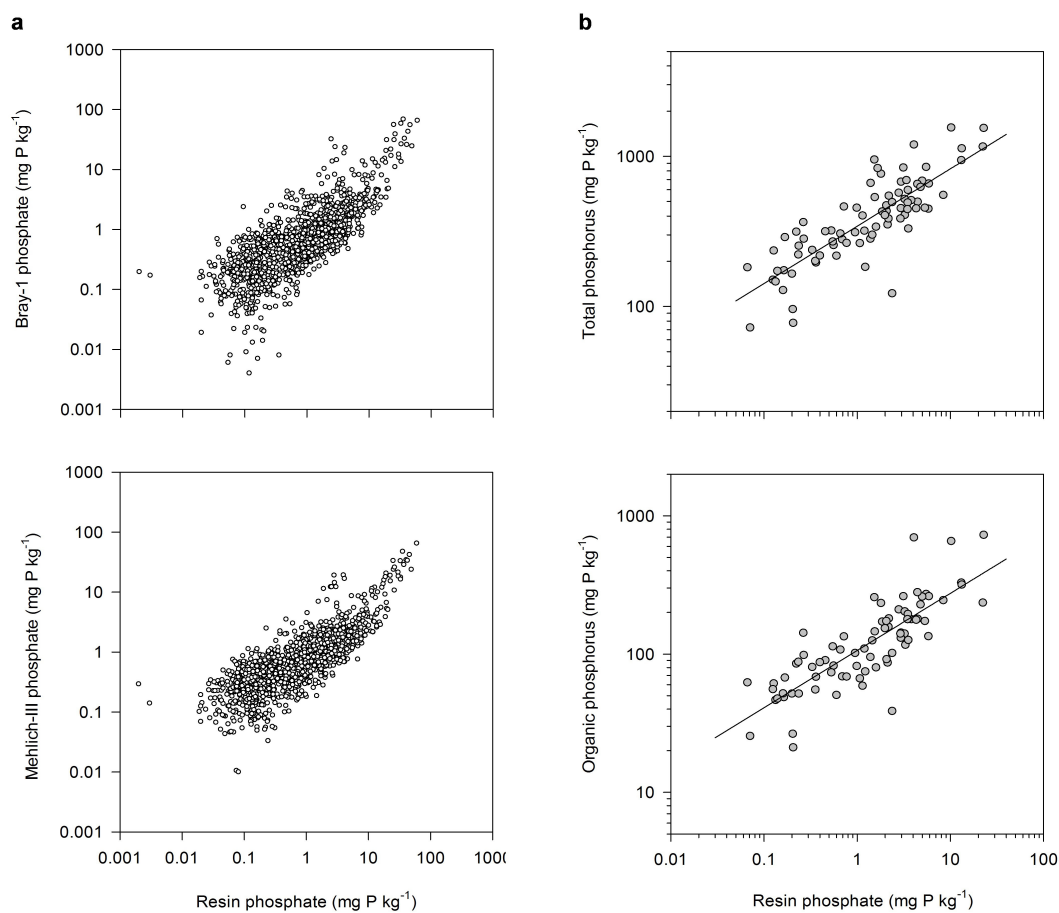
a, Modelled responses of common species to resin phosphate as adult trees and saplings. Trees were defined as being 100 mm dbh and saplings were defined as being 10 mm dbh. The responses of trees and saplings are unrelated by simple linear regression ($R^2 = 0.006$; $P = 0.74$). As trees 90% of species have a positive response to increasing resin phosphate concentrations (points above the horizontal dotted line), and as saplings 84% of species have a positive response to increasing resin phosphate concentrations (points to the right of the vertical dotted line). Only three common species responded negatively as both small and large trees. **b**, Piecewise linear regression model using common widespread species, showing the relationship fitted to the response of growth (log-transformed) to resin phosphate concentration for trees >100 mm dbh (top) and saplings <100 mm dbh (bottom). The black line is the community-wide mean, or fixed response. Each grey line is the fit for one species and blue dots are the growth rates of individual trees. For trees, the break point between large and small responses to phosphorus is at 1.6 mg P kg^{-1} resin phosphate (red dashed vertical line; 95% credible interval 1.3–2.0). To the left of this break, $s_1 = 0.16$ (95% credible interval 0.06–0.28) and to the right, $s_2 = 0.01$ (–0.01–0.03). For saplings, s_2 was

significantly positive. However, the two slopes had widely overlapping credible intervals, forcing us to accept the null hypothesis of no change in slope. **c**, Specific phosphatase activity and resin phosphate for 83 sites under lowland tropical forest in Panama, showing phosphomonoesterase activity and phosphodiesterase activity expressed on the basis of the soil microbial biomass carbon (left) and total soil organic carbon (right). For both transformations, the relationships are almost identical to those for non-standardized activities, but the models explain a slightly smaller proportion of the variance. The hydrolysis product is methylumbelliferone and model fits are exponential functions determined by nonlinear regression. **d**, The proportion of the widespread species at a site that have negative or positive associations with soil phosphorus, against the resin phosphate concentration for 72 lowland tropical forests in Panama. Species with negative associations with soil phosphorus (low-phosphorus affinity), open blue circles and blue line; species with positive associations with soil phosphorus (high-phosphorus affinity), red circles and red line. The point at which the proportion of low-affinity species equals the proportion of high-affinity species corresponds to a resin phosphate concentration of $2.18 \text{ mg P kg}^{-1}$.



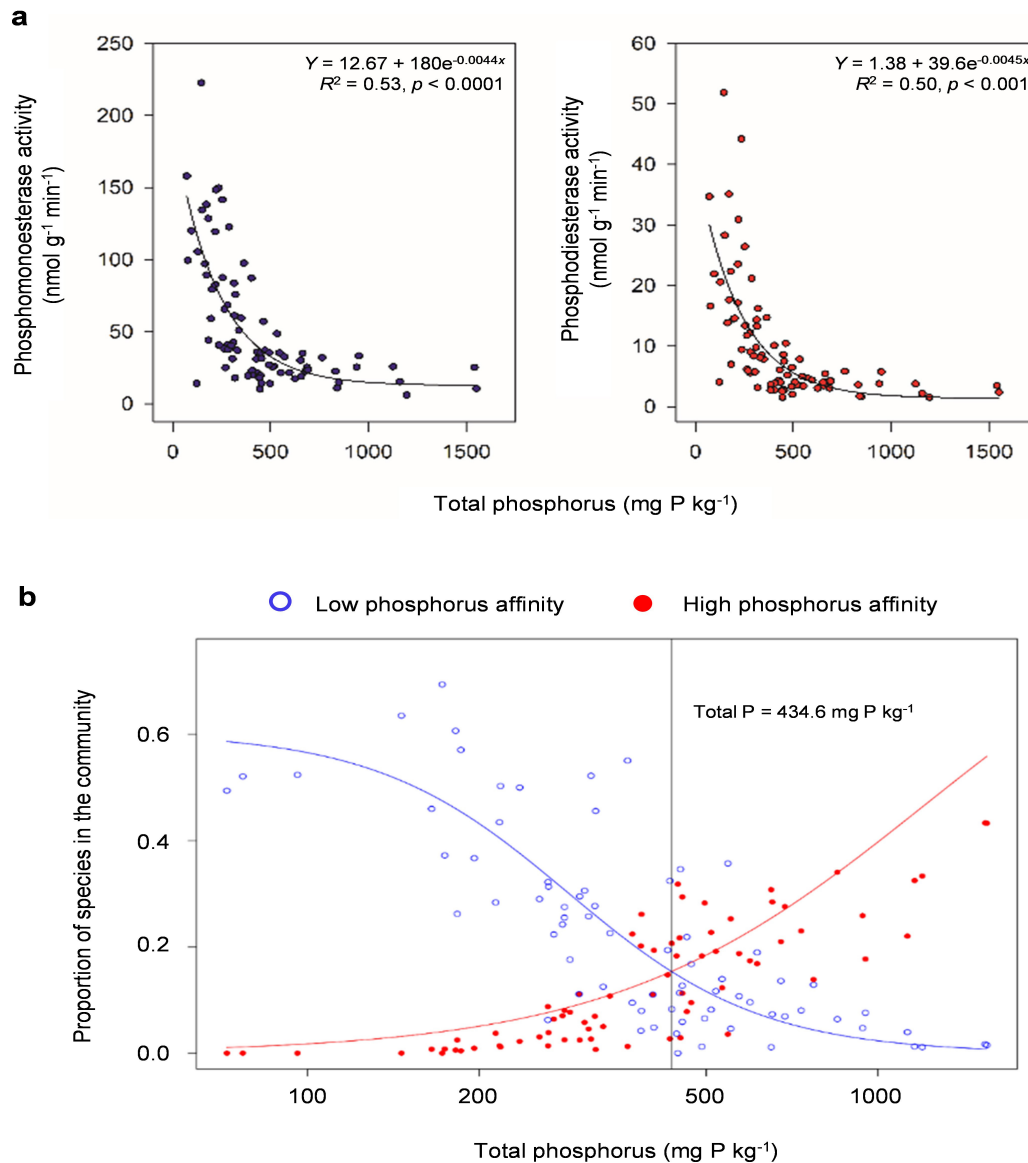
Extended Data Figure 3 | Growth responses to phosphorus at the species and community levels. **a**, Similarity in the growth rates of individual common species as predicted by the hierarchical model at three different resin phosphate concentrations. Each point represents the growth rate of a single species as estimated from the model, assuming intermediate moisture and a tree of 100 mm dbh. The graphs show the predicted species responses at intermediate resin phosphate (x axis, predicted growth – midP; as shown in Fig. 2a) against the predicted responses at low resin phosphate (predicted growth – lowP; left) and high resin phosphate (predicted growth – highP; right) concentrations. Only species that are common in the dataset (growth data available for >20 individuals) are plotted. The relative estimated responses are virtually identical across the entire phosphorus gradient. **b**, Observed growth rates as a function of species phosphorus affinities, with growth rates of individual trees ≥ 100 mm dbh shown in black and species-level median growth for the 362 species with estimated phosphorus affinities in blue. The y axis (growth) is log-transformed. The blue line shows a standard

linear regression between log-transformed growth and phosphorus affinity (effect size), using median species growth ($n = 362$), weighted by species abundance. The slope (-0.13) is significantly different from zero ($P = 0.00014$), demonstrating that growth rates were greater for species with low-phosphorus affinity. **c**, Standing above-ground biomass (AGB) (top) and annual relative AGB growth (that is, standardized by the total AGB) (bottom) as a function of resin phosphate concentration. Data are from 32 plots across the phosphorus gradient in Panama. The resin phosphate scale is logarithmic. The linear regression relating standing AGB to log(resin phosphate), dry-season intensity and successional state revealed a slight negative but non-significant effect of phosphorus on biomass (slope = -7.9 , $P = 0.37$). The same regression for relative AGB growth was likewise negative but not significant (slope = -0.002 , $P = 0.09$). Biomass was significantly and negatively related to dry-season intensity (that is, more biomass at wetter sites) ($P = 0.003$), but relative biomass growth was not correlated with dry season intensity ($P = 0.07$).



Extended Data Figure 4 | Relationships between resin phosphate and other measures of soil phosphorus. **a**, Comparison of resin phosphate concentration and two common extraction procedures for plant-available phosphorus, showing values for 1,184 fresh (field-moist) soil samples at depths of up to 100 cm in lowland tropical forests of Panama. Relationships are shown for Bray-1 phosphate (top) and Mehlich-III phosphate (bottom). For both extractions, phosphate was determined in the extracts by automated molybdate colorimetry. Resin phosphate is strongly correlated to Bray-1 phosphate (Pearson product-moment

correlation 0.81, $P < 0.0001$) and Mehlich-III phosphate (Pearson product-moment correlation 0.87, $P < 0.0001$). **b**, Relationship between resin phosphate concentration and total phosphorus (top) and organic phosphorus (bottom). Data are from soils from 83 sites in central Panama, with each value being the mean of multiple individual soil samples at a single site. The relationships are described by the following equations, derived from linear regression of log-transformed data: total phosphorus: $y = 342.43 \times (x^{0.3821})$, $R^2 = 0.68$, $P < 0.001$; organic phosphorus: $y = 106.01 \times (x^{0.4136})$, $R^2 = 0.66$, $P < 0.001$.



Extended Data Figure 5 | Phosphorus limitation threshold for total phosphorus. **a**, Relationships between phosphatase activities and total phosphorus concentrations in soils from 83 sites under lowland tropical forest in Panama. The figure shows phosphomonoesterase activity (blue circles, left) and phosphodiesterase activity (red circles, right). The hydrolysis product is methylumbelliferone and the model fits are negative exponential functions determined by nonlinear regression. The activity of both phosphatases decreases markedly at total phosphorus concentrations >400 mg P kg⁻¹. **b**, The proportion of species at a site that has a negative

association with soil phosphorus (low-phosphorus affinity, blue circles and blue line) or positive association with soil phosphorus (high-phosphorus affinity, red circles and red line) against total soil phosphorus for 72 lowland tropical forests in Panama. The models are sigmoidal fits and phosphorus associations are defined as effect sizes >0.8 (positive affinity) or <-0.8 (negative affinity). The point at which the proportion of low-affinity species exceeds the proportion of high-affinity species corresponds to a total phosphorus concentration of 435 mg P kg⁻¹.

Extended Data Table 1 | Results of the full model used for analysis of the effect of resin on tree growth

Fixed effects	Estimate	Std. Error	<i>t</i> value	Probability (<i>P</i>)
Intercept	−0.1459	0.0414	−3.52	
Log(dbh)	0.5719	0.0246	23.23	< 0.0001
Log(dbh) ²	−0.0667	0.0153	−4.36	< 0.0001
Moisture deficit	0.0100	0.0177	0.57	0.003
Log(dbh) × moisture deficit	−0.0505	0.0197	−2.57	0.002
Log(Resin P)	0.0716	0.0203	3.52	< 0.0001
Log(dbh) × Log(Resin P)	0.0059	0.0150	0.40	0.72

Random effects	Parameter	Variance	Std. Dev.	Probability (<i>P</i>)
Species	Intercept	0.410	0.640	
	Log(dbh)	0.077	0.277	< 0.0001
	Moisture deficit	0.015	0.123	0.002
	Log(dbh) ²	0.027	0.164	< 0.0001
	Log(dbh) × deficit	0.023	0.151	< 0.0001
	Log(Resin P)	0.031	0.176	0.0015
	Log(dbh) × Log(Resin P)	0.007	0.085	0.72
Residual		1.397	1.182	

P values indicate significant improvements in the model when each individual parameter was included; they were calculated using LogLik model comparisons after dropping one fixed effect parameter at a time (relative to a model in which only the intercept varied randomly per species). Species random effects show the variance among species in the response to each fixed effect value. *P* values for random effects were generated from separate model runs shown in Extended Data Table 2. Response variable = log(growth); number of observations = 18,970; number of species = 541. dbh, diameter at breast height.

Extended Data Table 2 | Models evaluating the effects of moisture and nutrients on tree growth

(a) Moisture model

Fixed effects	Estimate	Std. Error	<i>t</i> value	Probability (<i>P</i>)
Intercept	−0.1535	0.0404	−3.80	
Log(dbh)	0.5804	0.0243	23.93	< 0.0001
Log(dbh) ²	−0.0647	0.0152	−4.26	< 0.0001
Moisture deficit	−0.0084	0.0177	−0.47	0.003
Log(dbh) × moisture deficit	−0.0606	0.0189	−3.21	0.001

Random effects	Name	Variance	Std. Dev.	Probability (<i>P</i>)
Species	Intercept	0.409	0.640	
	Log(dbh)	0.078	0.279	< 0.0001
	Moisture deficit	0.019	0.137	0.002
	Log(dbh) ²	0.027	0.165	< 0.0001
	Log(dbh) × deficit	0.019	0.137	< 0.0001
Residual		1.408	1.187	

(b) Effects of adding individual nutrients to the model

	Estimate	Std. Error	<i>t</i> value	Fixed effects <i>P</i> value	Random effects <i>P</i> value
Log(Resin P)	0.0707	0.0181	3.91	< 0.0001	0.0015
Log(Total Inorganic N)	0.0152	0.0096	1.59	0.11	ns
Log(Mehlich Ca)	0.0522	0.0169	3.09	< 0.0001	< 0.0001
Log(Mehlich K)	0.0044	0.0115	0.38	0.77	ns
Log(Mehlich Mn)	0.0046	0.015	0.49	0.63	ns

a. Model evaluating the effect of moisture deficit and dbh on growth, with no nutrients included. All fixed and random-effect parameters included in the moisture model significantly improved the model based on AIC model comparisons. **b.** Parameters and *P* values obtained from adding a single nutrient parameter at a time to the moisture model. *P* values evaluate the improvement in the model based on LogLik model comparisons after adding each nutrient parameter, one at a time, first as a fixed effect and then to the species random effect. Response variable = log(growth); number of observations = 18,970; number of species = 541. ns, not significant (*P* > 0.1).

Extended Data Table 3 | Model evaluating the effect on tree growth of resin phosphate and Mehlich calcium

Fixed effects	Estimate	Std. Error	<i>t</i> value	Probability (<i>P</i>)
Intercept	−0.134	0.041	−3.27	
Log(dbh)	0.574	0.024	23.74	*
Log(dbh) ²	−0.067	0.015	−4.43	*
Moisture deficit	−0.005	0.019	−0.25	*
Log(dbh) × moisture deficit	−0.049	0.019	−2.60	*
Log(Resin P)	0.052	0.020	2.65	0.008 †
Log(Mehlich Ca)	0.027	0.017	1.63	0.11 ‡

P values were calculated using LogLik model comparisons after adding the nutrient parameters, one at a time, to the moisture model in Extended Data Table 2. Response variable = log(growth); number of observations = 18,970; number of species = 541.

*See Extended Data Table 2 for *P* values.

†Also significant (*P* = 0.02) when log(resin P) was added to the Mehlich Ca model in Extended Data Table 2.

‡Also non-significant (*P* = 1.0) when log(Mehlich Ca) was added to the resin P model in Extended Data Table 2.

Extended Data Table 4 | Additional model runs to examine the influence of plot-level parameters on tree growth response to phosphorus**(a) Model including a plot random effect**

Fixed effects	Estimate	Std. Error	<i>t</i> value
Intercept	−0.1330	0.0531	−2.51
Log(dbh)	0.5698	0.0244	23.39
Log(dbh) ²	−0.0716	0.0152	−4.73
Moisture deficit	−0.0100	0.0403	−0.25
Log(Resin P)	0.0942	0.0409	2.30
Log(dbh) × moisture deficit	−0.0290	0.0198	−1.47
Log(dbh) × log(Resin P)	0.0091	0.0154	0.59

(b) Model including stem density

Fixed effects	Estimate	Std. Error	<i>t</i> value
Intercept	−0.1413	0.0412	−3.43
Log(dbh)	0.5826	0.0243	24.00
Log(dbh) ²	−0.0710	0.0154	−4.62
Moisture deficit	0.0356	0.0181	1.96
Log(Resin P)	0.0514	0.0189	2.72
Log(dbh) × moisture deficit	−0.0517	0.0191	−2.71
Log(dbh) × log(Resin P)	0.0125	0.0135	0.93
Stem density	−0.0803	0.0143	−5.63

a. Model including a plot random effect. **b.** Model including the effect of stem density. Both models are for trees ≥ 10 mm dbh. Stem density was calculated as the number of stems ≥ 100 mm diameter. *P* values were not calculated, but *t*-values < -2 or > 2 are generally significant at $P > 0.05$ for sample sizes > 60 . Response variable = log(growth); number of observations = 18,970; number of species = 541.

Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours

Xiaotu Ma^{1*}, Yu Liu^{1*}, Yanling Liu¹, Ludmil B. Alexandrov², Michael N. Edmonson¹, Charles Gawad¹, Xin Zhou¹, Yongjin Li¹, Michael C. Rusch¹, John Easton¹, Robert Huether^{3†}, Veronica Gonzalez-Pena⁴, Mark R. Wilkinson¹, Leandro C. Hermida⁵, Sean Davis⁶, Edgar Sioson¹, Stanley Pounds⁷, Xueyuan Cao⁷, Rhonda E. Ries⁸, Zhaoming Wang¹, Xiang Chen¹, Li Dong¹, Sharon J. Diskin⁹, Malcolm A. Smith¹⁰, Jaime M. Guidry Auvil⁵, Paul S. Meltzer⁶, Ching C. Lau^{11,12}, Elizabeth J. Perlman¹³, John M. Maris⁹, Soheil Meshinchi⁸, Stephen P. Hunger⁹, Daniela S. Gerhard⁵ & Jinghui Zhang¹

Analysis of molecular aberrations across multiple cancer types, known as pan-cancer analysis, identifies commonalities and differences in key biological processes that are dysregulated in cancer cells from diverse lineages. Pan-cancer analyses have been performed for adult^{1–4} but not paediatric cancers, which commonly occur in developing mesodermic rather than adult epithelial tissues⁵. Here we present a pan-cancer study of somatic alterations, including single nucleotide variants, small insertions or deletions, structural variations, copy number alterations, gene fusions and internal tandem duplications in 1,699 paediatric leukaemias and solid tumours across six histotypes, with whole-genome, whole-exome and transcriptome sequencing data processed under a uniform analytical framework. We report 142 driver genes in paediatric cancers, of which only 45% match those found in adult pan-cancer studies; copy number alterations and structural variants constituted the majority (62%) of events. Eleven genome-wide mutational signatures were identified, including one attributed to ultraviolet-light exposure in eight aneuploid leukaemias. Transcription of the mutant allele was detectable for 34% of protein-coding mutations, and 20% exhibited allele-specific expression. These data provide a comprehensive genomic architecture for paediatric cancers and emphasize the need for paediatric cancer-specific development of precision therapies.

Paired tumour and normal samples from 1,699 patients with paediatric cancers enrolled in Children's Oncology Group clinical trials were analysed, including 689 B-lineage acute lymphoblastic leukaemias (B-ALL), 267 T-lineage ALLs (T-ALL), 210 acute myeloid leukaemias (AML), 316 neuroblastomas (NBL), 128 Wilms tumours and 89 osteosarcomas (Extended Data Fig. 1a–c). All tumour specimens were obtained at initial diagnosis, and 98.5% of patients were 20 years of age or younger (see Methods, Extended Data Fig. 1d).

The median somatic mutation rate ranged from 0.17 per million bases (Mb) in AML and Wilms tumours to 0.79 in osteosarcomas (Fig. 1a, b), lower than the 1–10 per Mb found in common adult cancers⁶. Genome-wide analysis (see Methods) identified 11 mutational signatures (T-1 through T-11; Fig. 1c–e and Supplementary Table 1a–c). Signatures T-1 through T-9 corresponded to known COSMIC signatures⁷, whereas T-10 and T-11 were novel but enriched in mutations with a low (<0.3) mutant allele fraction (MAF).

Signatures T-1 and T-4 (clock-like endogenous mutational processes) were present in all samples and contributed to large proportions of all mutations in T-ALL (97%), AML (63%), B-ALL (36%), and Wilms tumours (28%). T-2 and T-7 (APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like)) were highly enriched in B-ALLs with *ETV6-RUNX1* fusions (15-fold and 9-fold enrichment for T-2 and T-7, respectively; Supplementary Table 1e). T-3 (homologous recombination deficiency) was present in many childhood cancers, including osteosarcomas (18 of 19), NBLs (59 of 137), Wilms tumours (28 of 81), and B-ALL (47 of 218). T-8 (8-oxoguanine DNA damage) was present in a small proportion (4.5–12%) of AML, B-ALL, osteosarcoma, and Wilms tumour samples. T-8 was also present in many (36%) NBL samples and was associated with age at diagnosis (Supplementary Table 1d). T-9 (DNA repair deficiency) was present in two B-ALLs, including one (sample PARJSR) with a somatic *MSH6* frameshift mutation. T-2, T-3, T-5, T-7, T-8, and T-9 were enriched among the 39 samples with elevated mutation rates in each histotype (Fig. 1d).

The T-5 ultraviolet-light (UV)-exposure signature was unexpectedly present in eight B-ALL samples (Extended Data Fig. 2a–c). Although its mutation rate in B-ALL, ranging from 0.06 to 0.72 per Mb, was 100-fold lower than the average rate in adult (15.8 per Mb)⁸ and paediatric (14.4 per Mb)⁹ skin cancer, T-5 exhibited other features associated with UV-related DNA damage. Specifically, CC>TT dinucleotide mutations were enriched 110-fold in these eight B-ALL samples when compared with other samples ($P = 1.07 \times 10^{-7}$), which is consistent with pyrimidine dimer formation. Moreover, transcriptional strand bias in T-5 indicated that photodimer formation contributed to cytosine damage. The validity of T-5 was further confirmed by analysis of the mutation clonality, cross-platform concordance, genomic distribution and mutation spectrum of each sample (see Methods, Extended Data Fig. 2d–i), indicating that UV exposure or other mutational processes^{10,11} may contribute to paediatric leukemogenesis. Notably, all T-5 B-ALLs had aneuploid genomes ($P = 3 \times 10^{-5}$; two-sided binomial test; cohort frequency 24%) without any oncogenic fusions.

By analysing the enrichment^{12,13} of somatic alterations within each histotype or the pan-cancer cohort (see Methods), we identified 142 significantly mutated driver genes (Fig. 2a, Supplementary Table 2, Extended Data Fig. 3a). Somatic alterations in *CDKN2A*, which were predominantly deletions, occurred at the highest frequency, affecting

¹Computational Biology, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. ²Department of Cellular and Molecular Medicine and Department of Bioengineering and Moores Cancer Center, University of California, San Diego, La Jolla, California 92093, USA. ³Independent Researcher, Chicago, Illinois 60654, USA. ⁴Oncology, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. ⁵Office of Cancer Genomics, National Cancer Institute, Bethesda, Maryland 20892, USA. ⁶Genetics Branch, Center for Cancer Research, National Cancer Institute, NIH, Bethesda, Maryland 20892, USA. ⁷Department of Biostatistics, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. ⁸Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. ⁹Division of Oncology and Center for Childhood Cancer Research, Children's Hospital of Philadelphia and the Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ¹⁰Cancer Therapy Evaluation Program, National Cancer Institute, Bethesda, Maryland 20892, USA. ¹¹Division of Hematology-Oncology, Connecticut Children's Medical Center, Hartford, Connecticut 06106, USA. ¹²The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut 06032, USA.

¹³Department of Pathology and Laboratory Medicine, Ann & Robert H. Lurie Children's Hospital of Chicago, Robert H. Lurie Cancer Center, Northwestern University, Chicago, Illinois 60208, USA.

[†]Present address: Tempus Labs Inc., Chicago, Illinois 60654, USA.

*These authors contributed equally to this work.

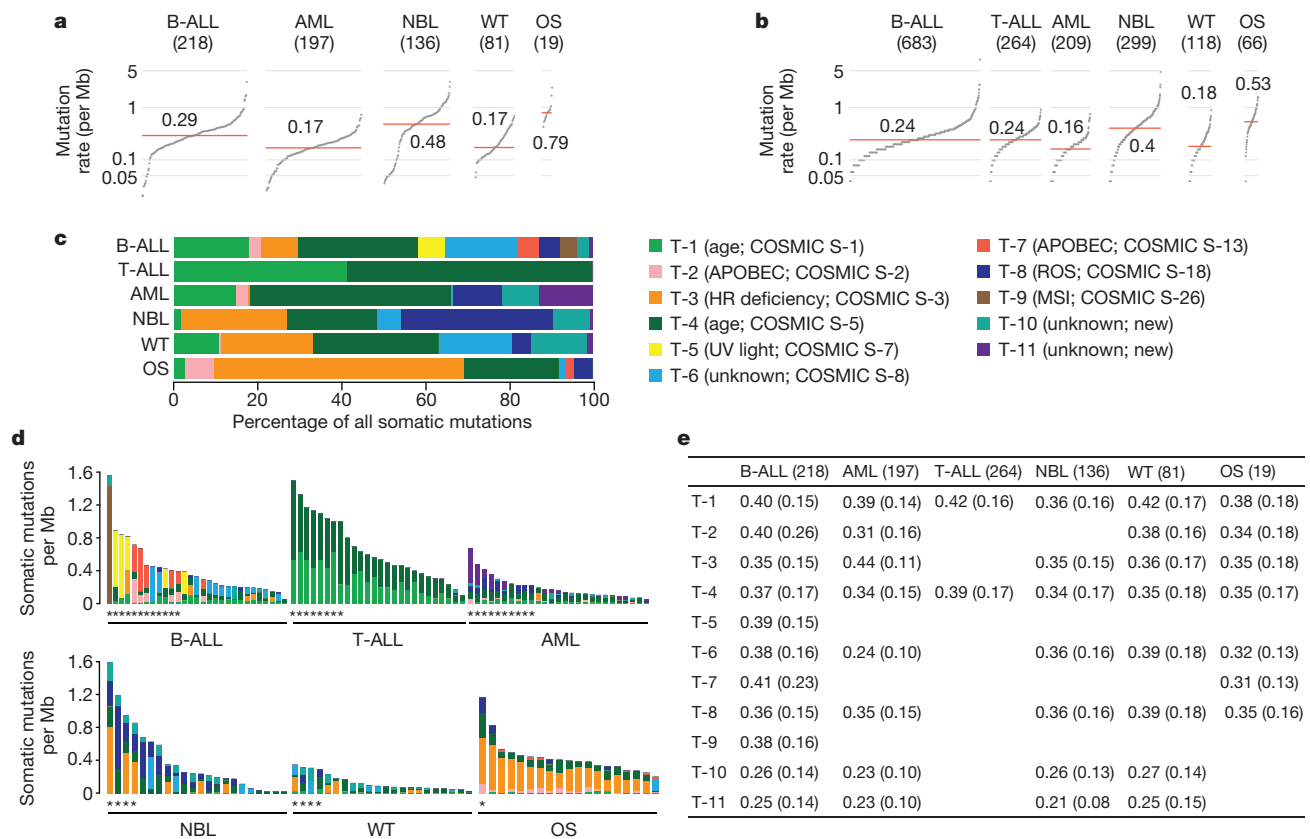


Figure 1 | Somatic mutation rate and signature. Sample size of each histotype is shown in parentheses. Mutation rate using non-coding SNVs from WGS (**a**) and coding SNVs from WGS and WES (**b**). Red line, median. **a** and **b** are scaled to the total number of samples with WGS ($n = 651$), WGS or WES ($n = 1,639$), respectively. **c**, Mutational signatures

207 of 267 (78%) T-ALLs, 91 of 218 (42%) B-ALLs and 2 of 19 (11%) osteosarcomas (Extended Data Fig. 3b). More than half (73) of the driver genes were specific to a single histotype, such as *TAL1* for T-ALL and *ALK* for NBL (Extended Data Fig. 3c). Genes that were mutated in both leukaemias and the three solid tumour histotypes accounted for only 17% of driver genes (Extended Data Fig. 3e), of which some genes had various types of somatic alteration. For example, *STAG2*, a known driver gene for Ewing's sarcoma¹⁴ and adult AML¹⁵, exhibited five different types of somatic alteration (single nucleotide variants (SNVs), small insertions or deletions (indels), copy number alterations (CNAs), structural variants and internal tandem duplications (ITDs)) across five histotypes (Extended Data Fig. 4a–d). Nine *STAG2* variants were predicted to cause protein truncation, including four predicted by aberrant transcripts in RNA sequencing (RNA-seq). Notably, 78 of 142 driver genes (Supplementary Table 2) were not found in adult pan-cancer studies^{1–4}, and 43 (Fig. 2a and Extended Data Fig. 3a) were not found in the Cancer Gene Census (v81)¹⁶. Thirty-seven were absent from both sources, although mutations in cancer have been reported for 29 of these genes, such as *NIPBL*^{17–19} and *LEMD3*²⁰ (Extended Data Fig. 4p, q). Nearly half (40–50%) of point mutations in leukaemia and NBL driver genes had low MAFs (<0.3), indicative of subclonal mutations contributing to tumorigenesis (Extended Data Fig. 3f).

Three hundred and four gene-pairs exhibited statistically significant ($P < 0.05$, two-sided Fisher's exact test; Fig. 2b, Supplementary Table 3) co-occurrence (for example, *USP7* and *TAL1* in T-ALL²¹) or mutual exclusivity (for example, *MYCN* and *ATRX* in NBL²²). The analysis also unveiled novel co-occurrences (for example, *ETV6* and *IKZF1* in AML and *CREBBP* and *EP300* in B-ALL) and mutual exclusivities (for example, *SHANK2* and *MYCN* in NBL and *PAX5* and *TP53* in B-ALL).

identified from WGS and T-ALL WES data and their contribution in each histotype. **d**, Mutation spectrum of representative samples in each histotype. Hypermutators (three s.d. above mean rate of corresponding histotype) are labelled with an asterisk. **e**, Mean and s.d. of MAF of each signature in each histotype.

Because of reduced power for detecting low-frequency drivers² (detection limits were 1% for the entire cohort and 3% for individual histotypes with more than 200 samples; Extended Data Fig. 5 and Methods), we performed subnetwork analyses³ and variant pathogenicity classification²³ (see Methods), identifying 184 variants in 82 additional genes (Supplementary Table 4 and Extended Data Fig. 4e, f). A notable example is the *MAP3K4* G1366R mutation, which was found in one T-ALL, two B-ALLs, and one Wilms tumour. *MAP3K4* is a member of the MAPK family²⁴ and structural modelling indicates that the G1366R mutation is likely to cause disruption of normal inhibitory domain binding and kinase dynamics²⁴ (Extended Data Fig. 4l, m). Several genes in which structural variants were found (*PDGFRA*, *CDK4*, *YAP1*, *UBTF*) are listed in Extended Data Fig. 4.

While the percentage of tumours with point mutations in driver genes was highly consistent between whole-genome sequencing (WGS) and whole-exome sequencing (WES) (Fig. 3a), WGS makes it possible to detect CNAs and structural variants, which are frequently driver events for paediatric cancers. For example, 72% of NBL tumours analysed by WGS had at least one driver variant compared to 26% of those analysed by WES (Fig. 3a and Extended Data Fig. 4j, k). Furthermore, integrative analyses of CNAs and structural variants with WGS data revealed chromothripsis (that is, massive rearrangements caused by a single catastrophic event) in 11% of all samples (13 in osteosarcomas, 15 in Wilms tumours, 22 in NBL, 14 in B-ALL, and 6 in AML; Extended Data Fig. 1f). We next performed pathway analyses (see Methods) on 654 samples analysed by WGS and 264 T-ALL samples analysed by both WES and single nucleotide polymorphism (SNP) arrays, totaling 682 leukaemias and 236 solid tumours.

The 21 biological pathways that were disrupted by driver alterations were either common (for example, cell cycle and epigenetic regulation)

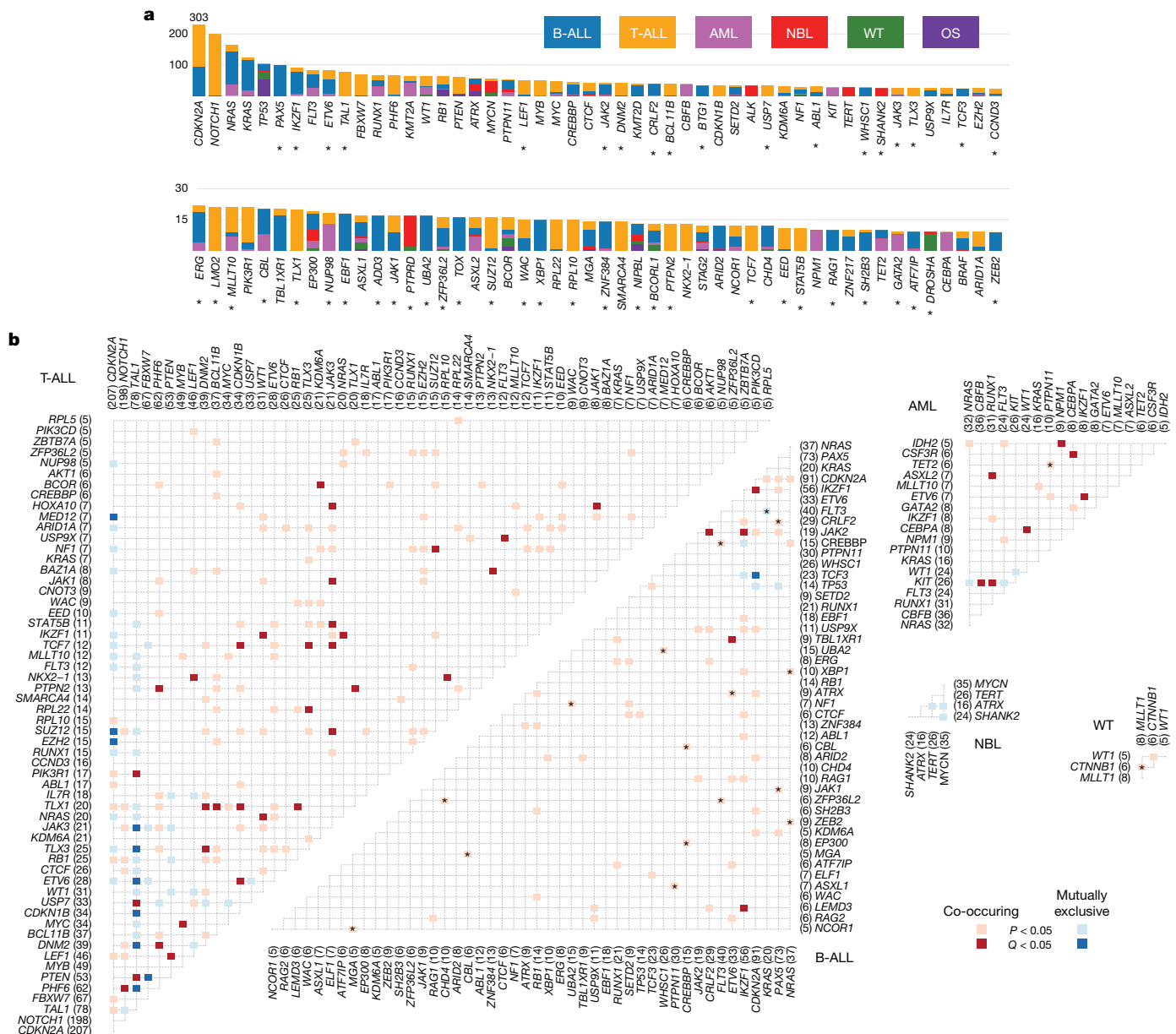


Figure 2 | Candidate driver genes in paediatric cancer. a, Top 100 recurrently mutated genes: case count for each histotype is shown in the same colour as the legend. Asterisk indicates gene not reported in prior adult pan-cancer analyses. **b**, Statistically significant pairwise relationships ($P < 0.05$; two-sided Fisher's exact test) for co-occurrence

or histotype-specific (for example, JAK–STAT, Wnt/ β -catenin, and NOTCH signalling) (Fig. 3b). More importantly, the genes that were mutated in each pathway differed between histotypes. One example is signalling pathways such as RAS, JAK–STAT and PI3K (Fig. 3c). For genes in these pathways, somatic alterations in solid tumours primarily occurred in *ALK*, *NF1*, and *PTEN*, whereas nearly all mutations in *FLT3*, *PIK3CA*, *PIK3R1*, and *RAS* were found in leukaemias. Although many biological processes are dysregulated in both paediatric and adult cancers^{1,2,4}, the affected genes may be either paediatric-specific (for example, transcription factors and JAK–STAT pathway genes) or common to both (for example, cell cycle genes and epigenetic modifiers). Notably, two novel *KRAS* isoforms were detected in 70% of leukaemias but rarely in solid tumours (Extended Data Fig. 6).

Evaluation of mutant allele expression makes it possible to assess the effects on the gene product and to detect potential epigenetic regulation

(red) or exclusivity (blue) in each histotype. Gene pairs with $Q < 0.05$ are coloured dark red (co-occurring) or dark blue (exclusive) to account for false discovery rate. Significance detected only in WGS + WES samples is marked with an asterisk. Shown in parentheses are number of mutated samples.

that may cause allelic imbalance. Here we present this analysis on 6,959 coding mutations with matching WGS and RNA-seq data. RNA-seq expression clusters confirmed the tissue of origin of each histotype (Extended Data Fig. 7). Mutant alleles were expressed for 34% of these mutations, which is consistent with previous reports^{25–27}. The expression of mutant alleles is generally associated with corresponding DNA MAF and the expression levels of host genes (Fig. 4a); however, exceptions can be found due to X-inactivation, imprinting, nonsense-mediated decay or complex structural re-arrangements (Extended Data Fig. 8a).

Allele-specific expression (ASE) was evaluated for 2,477 somatic point mutations with sufficient read-depth in DNA and RNA-seq (see Methods). Of 486 candidate ASE mutations (Supplementary Table 5), 279 had no detectable expression of the mutant allele, and a comparable DNA MAF distribution was found for truncating and non-truncating mutations ($P = 0.5$, two-sided Wilcoxon rank-sum test, Extended Data

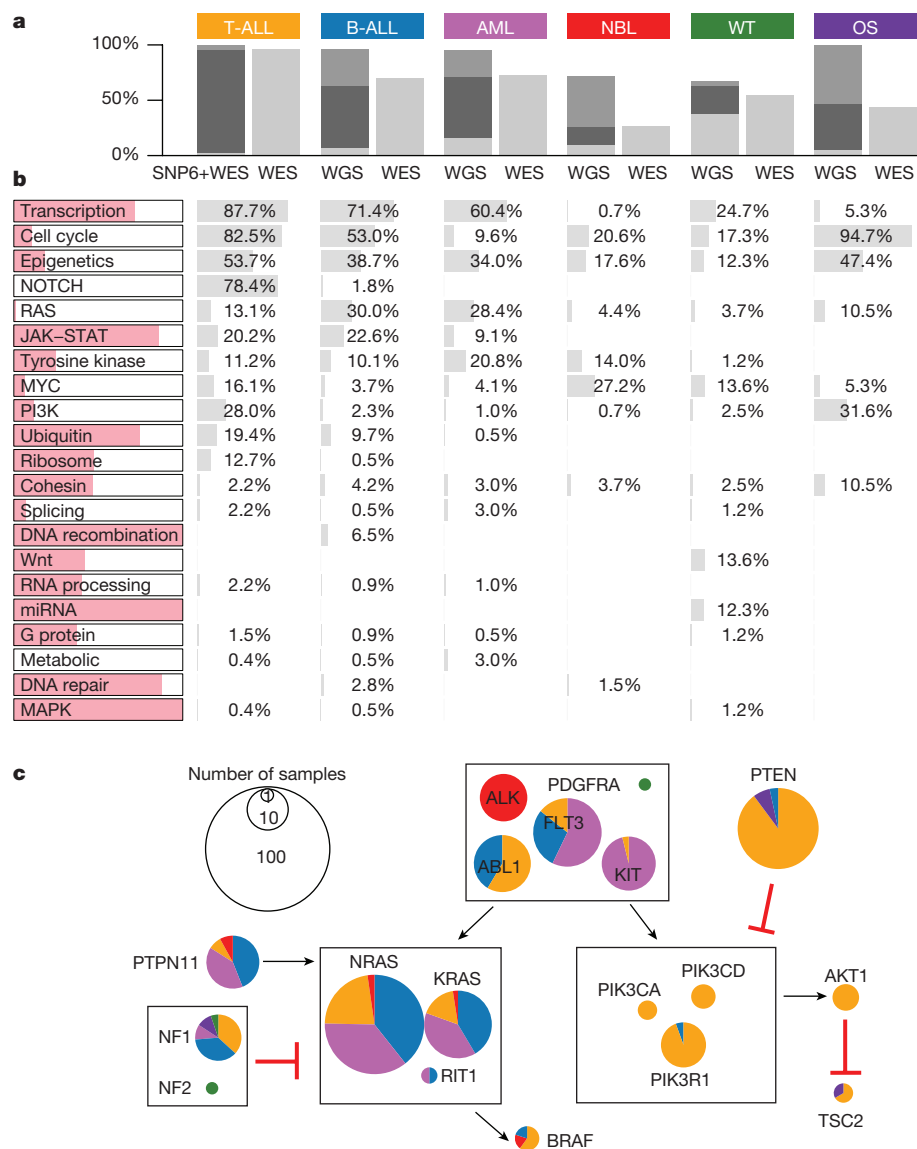


Figure 3 | Biological processes with somatic alterations in paediatric cancer. **a**, Percentage of tumours with at least one driver alteration are shown for each histotype. WGS-analysed tumours may have point mutations (light grey), CNAs or structural variations (SV) (dark grey), or both (black). For T-ALL, CNAs were derived from SNP array. **b**, Percentage of tumours within each histotype that have somatic

alterations in 21 biological pathways; histotype ordering is as in **a**. The coloured portion of each pathway indicates the percentage of variants in genes that are absent in three TCGA pan-cancer studies. **c**, Mutation occurrence by histotype in RAS, tyrosine kinase, and PI3K pathways.

Fig. 8b). Of the remaining 207 candidate ASE mutations, 76% of truncating mutations exhibited suppression of the mutant allele ($P = 7 \times 10^{-5}$; two-sided binomial test), while 87% of hotspot mutations showed the opposite trend of elevated expression ($P = 6 \times 10^{-5}$; two-sided binomial test; Fig. 4b, Extended Data Fig. 8c). Excluding hotspot mutations resulted in equal distribution of suppression versus elevation (66 versus 55) for the remaining 121 non-truncating ASE mutations ($P = 0.4$; two-sided binomial test).

Subclonal loss-of-heterozygosity (LOH) in tumours is a confounding factor for ASE analysis. For example, significant allelic imbalance between tumour DNA and RNA MAF of *WT1* D447N in an AML that also harboured a subclonal 11p copy-neutral LOH (Fig. 4c) could be attributed to ASE or *WT1* expression of a subclone with a double-hit of D447N mutation and 11p LOH. To address this, we performed single-cell DNA sequencing on 63 germline variants on 11p and the somatic point mutations. We confirmed ASE by establishing that *WT1* D447N and 11p LOH occurred in separate subclones (Fig. 4c and Extended Data Fig. 9a, b). The resulting genotype data projected that

one *WT1* allele was silenced in a common ancestor and the other was lost in the three descendant subclones by 11p LOH, acquisition of the *WT1* D447N mutation, or focal deletion. Two additional AMLs with *WT1* D447N also exhibited ASE (Extended Data Fig. 9c), implying that loss of *WT1* expression by epigenetic silencing or mutations in *cis*-regulatory elements is not rare in AML. Similarly, single-cell sequencing of an ALL sample confirmed ASE of a *JAK2* hotspot mutation (Extended Data Fig. 9d).

The somatic variants used for this study are available at the National Cancer Institute TARGET Data Matrix and our ProteinPaint²⁸ portal, which provides an interactive heat map viewer for exploring mutations, genes, and pathways across the six histotypes (Extended Data Fig. 10). The portal also hosts the somatic variants analysed by the companion paediatric pan-cancer study of 961 tumours from 24 histotypes, including 559 central nervous system tumours²⁹. We anticipate that these complementary pan-cancer datasets will be an important resource for investigations of functional validation and implementation of clinical genomics for paediatric cancers.

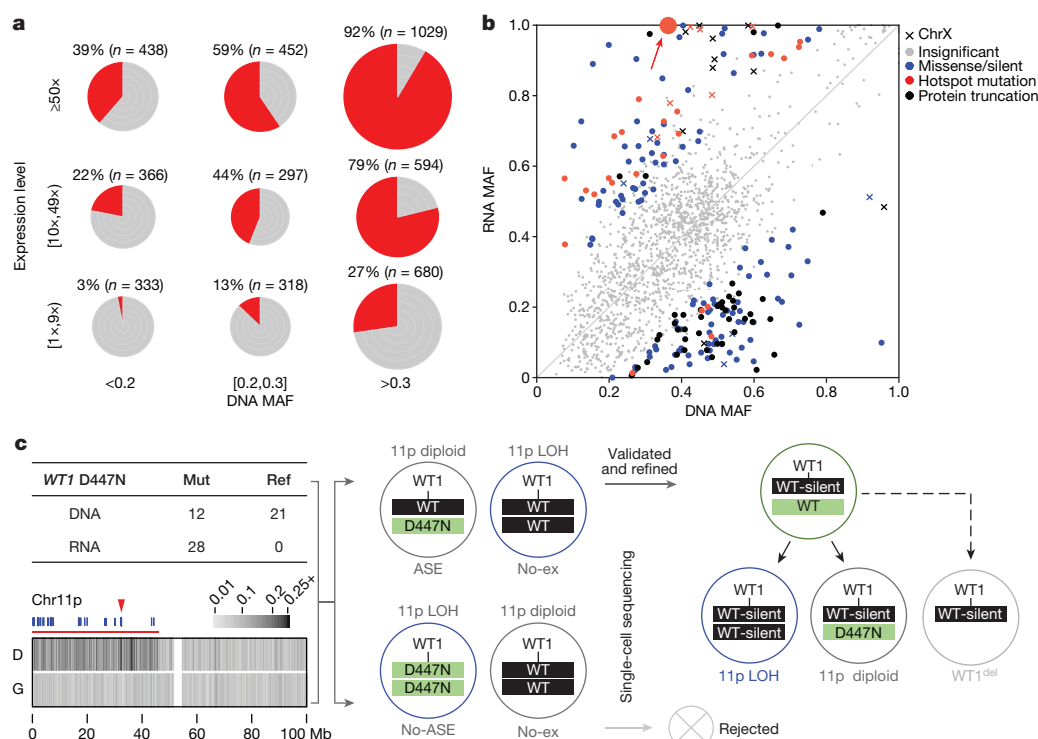


Figure 4 | Mutant allele expression. **a**, Percentage of expressed mutations (red) categorized by DNA MAF (x axis) and expression level (y axis). Circle size is proportional to mutation counts. **b**, Detection of ASE in expressed mutations by comparing DNA and RNA MAF in 443 samples (solid colours, statistically significant (two-sided Fisher's exact test $Q < 0.01$ and effect size > 0.2); grey, not significant). **c**, Confirming ASE

for *WT1* D447N (red arrow in **b**) by single-cell sequencing. Presence of subclonal 11p LOH leads to two possible outcomes: the mutant allele is in either non-LOH subclone (top) or LOH subclone (bottom): the former suggests ASE and the latter rejects ASE due to homozygosity. No-ex: *WT1* not expressed.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 February 2017; accepted 17 January 2018.

Published online 28 February 2018.

- Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339 (2013).
- Lawrence, M. S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501 (2014).
- Leiserson, M. D. *et al.* Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nat. Genet.* **47**, 106–114 (2015).
- Zack, T. I. *et al.* Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **45**, 1134–1140 (2013).
- Downing, J. R. *et al.* The Pediatric Cancer Genome Project. *Nat. Genet.* **44**, 619–622 (2012).
- Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
- Alexandrov, L. B. Understanding the origins of human cancer. *Science* **350**, 1175–1177 (2015).
- Hayward, N. K. *et al.* Whole-genome landscapes of major melanoma subtypes. *Nature* **545**, 175–180 (2017).
- Lu, C. *et al.* The genomic landscape of childhood and adolescent melanoma. *J. Invest. Dermatol.* **135**, 816–823 (2015).
- Reid, T. M. & Loeb, L. A. Tandem double CC→TT mutations are produced by reactive oxygen species. *Proc. Natl Acad. Sci. USA* **90**, 3904–3907 (1993).
- Newcomb, T. G., Allen, K. J., Ktshelashvili, L. & Loeb, L. A. Detection of tandem CC→TT mutations induced by oxygen radicals using mutation-specific PCR. *Mutat. Res.* **427**, 21–30 (1999).
- Pounds, S. *et al.* A genomic random interval model for statistical analysis of genomic lesion data. *Bioinformatics* **29**, 2088–2095 (2013).
- Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
- Tirode, F. *et al.* Genomic landscape of Ewing sarcoma defines an aggressive subtype with co-association of STAG2 and TP53 mutations. *Cancer Discov.* **4**, 1342–1353 (2014).
- Ley, T. J. *et al.* Genomic and epigenomic landscapes of adult *de novo* acute myeloid leukemia. *N. Engl. J. Med.* **368**, 2059–2074 (2013).
- Futreal, P. A. *et al.* A census of human cancer genes. *Nat. Rev. Cancer* **4**, 177–183 (2004).
- Krantz, I. D. *et al.* Cornelia de Lange syndrome is caused by mutations in *NIPBL*, the human homolog of *Drosophila melanogaster* Nipped-B. *Nat. Genet.* **36**, 631–635 (2004).
- Tonkin, E. T., Wang, T. J., Lisgo, S., Bamshad, M. J. & Strachan, T. *NIPBL*, encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins and fly Nipped-B, is mutated in Cornelia de Lange syndrome. *Nat. Genet.* **36**, 636–641 (2004).
- Barber, T. D. *et al.* Chromatid cohesion defects may underlie chromosome instability in human colorectal cancers. *Proc. Natl Acad. Sci. USA* **105**, 3443–3448 (2008).
- Hellemans, J. *et al.* Loss-of-function mutations in *LEMD3* result in osteopoikilosis, Buschke–Ollendorff syndrome and melorheostosis. *Nat. Genet.* **36**, 1213–1218 (2004).
- Liu, Y. *et al.* The genomic landscape of pediatric and young adult T-lineage acute lymphoblastic leukemia. *Nat. Genet.* **49**, 1211–1218 (2017).
- Cheung, N. K. & Dyer, M. A. Neuroblastoma: developmental biology, cancer genomics and immunotherapy. *Nat. Rev. Cancer* **13**, 397–411 (2013).
- Zhang, J. *et al.* Germline mutations in predisposition genes in pediatric cancer. *N. Engl. J. Med.* **373**, 2336–2346 (2015).
- Mita, H., Tsutsui, J., Takekawa, M., Witten, E. A. & Saito, H. Regulation of MTK1/MEKK4 kinase activity by its N-terminal autoinhibitory domain and GADD45 binding. *Mol. Cell. Biol.* **22**, 4544–4555 (2002).
- Rashid, N. U. *et al.* Differential and limited expression of mutant alleles in multiple myeloma. *Blood* **124**, 3110–3117 (2014).
- Shah, S. P. *et al.* The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* **486**, 395–399 (2012).
- Govindan, R. *et al.* Genomic landscape of non-small cell lung cancer in smokers and never-smokers. *Cell* **150**, 1121–1134 (2012).
- Zhou, X. *et al.* Exploring genomic alteration in pediatric cancer using ProteinPaint. *Nat. Genet.* **48**, 4–6 (2016).
- Gröbner, S. N. *et al.* The landscape of genomic alterations across childhood cancers. *Nature* <http://doi.org/10.1038/nature25480> (2018).

Supplementary Information is available in the online version of the paper.

Acknowledgements This project was funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. HHSN261200800001E to J.Z. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial

products, or organizations imply endorsement by the US Government. This project was also funded by a supplement to the Children's Oncology Group Chair's grant CA098543 and by funding from ALSAC to St Jude Children's Research Hospital and by a Cancer Center Support (Core) grant (CA21765) from the National Cancer Institute to St Jude Children's Research Hospital. S.P.H. is the Jeffrey E. Perelman Distinguished Chair in the Department of Pediatrics at the Children's Hospital of Philadelphia. We thank B. Raphael and M. Reyna for their help on HotNet2 analysis; K. Shannon for discussion on novel KRAS isoforms; L. Robison and C. Howell for providing SEER incidence data for childhood ALL; S. Newman for assistance in CDKN2A deletion analysis; P. Gupta for data download support; T. Pugh, P. Gesuwan, T. Davidsen, C. G. Mullighan, J. Farrar, V. Huff, and S. Gadd for their participation in the TARGET Analysis Working Group and contributions to data generation and analysis; and L. Tian, S. Wang and N. Badders for assistance in revising the manuscript.

Author Contributions J.Z., D.S.G., and L.B.A. designed all experiments. J.Z., X.M., Yu.L., Ya.L., and L.B.A. performed all experiments and analyses with the help of M.N.E., Y.L., X.Z., E.S., M.C.R., S.P., X.Ca., L.C.H., M.R.W., S.D., R.E.R., Z.W., X.Ch., L.D. and J.M.G.A. The structural modelling was performed by R.H. The single-cell assays were performed by C.G., V. G.-P., and J.E. WGS, WES, and RNA-seq data were provided by P.S.M., C.C.L., E.J.P., S.J.D., J.M.M., S.M., and S.P.H. S.M. and S.P.H. provided the leukaemia specimen for single-cell sequencing. D.S.G.,

J.M.G.A., M.A.S., and L.C.H. oversaw the administrative and data management aspects of the TARGET project. The manuscript was written by J.Z., X.M., Y.L., L.B.A., J.M.M., and S.P.H. and was reviewed and edited by all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to J.Z. (jinghui.zhang@stjude.org).

Reviewer Information *Nature* thanks M. Meyerson and the other anonymous reviewer(s) for their contribution to the peer review of this work.



This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) licence. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons licence, users will need to obtain permission from the licence holder to reproduce the material. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

METHODS

Patient samples. Specimens were obtained through collaborations with the Children's Oncology Group (COG) and the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) project. Institutional review boards from the following institutions were responsible for oversight: Ann & Robert H. Lurie Children's Hospital, Fred Hutchinson Cancer Research Center, National Cancer Institute, St Jude's Children's Research Hospital, The Children's Hospital of Philadelphia, The University of New Mexico, Texas Children's Hospital, and The Hospital for Sick Children. In our cohort, osteosarcoma has a higher percentage of older patients because the age of onset has a bimodal distribution: the first peak occurs among adolescents and young adults, and the second (associated with Paget disease and with a different underlying biology³⁰) occurs among the elderly. We used an age cutoff of 40 years, which is typical for COG-conducted osteosarcoma trials³¹. Informed consent was obtained from all subjects.

Genomic datasets. WGS, WES, and RNA-seq data were downloaded from dbGaP with study identifier phs000218 (including phs000463, phs000464, phs000465, phs000467, phs000471, and phs000468). Among the 1,699 cases analysed, 45 B-ALLs^{32,33}, 197 AMLs³⁴, 264 T-ALLs²¹, 240 NBLs³⁵ and 115 Wilms tumours³⁶ have been included in published studies of individual histotypes. No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

WGS data analysis. WGS data were generated with Complete Genomics Inc. (CGI) technology with an average genome-wide coverage of 50× using 31- to 35-bp mate-paired reads, which was powered for detecting mutations in 94% of mappable exonic regions^{37,38}. Read pairs were mapped to hg19/GRCh37, and somatic SNVs, indels, and structural variants were analysed by comparing paired tumour and normal genomes using the CGI Cancer Sequencing service pipeline version 2^{38,39}.

For each case, we downloaded CGI-generated WGS files for somatic SNVs, indels, structural variants, and CNAs from the TARGET Data Matrix as the starting point for our analysis.

Filtering of point mutations. Putative somatic point mutations including SNVs and indels were extracted from Mutation Annotation Format files and run through a filter to remove false-positive calls. First, germline variants were filtered by using: (1) NHLBI Exome Sequencing Project (<http://evs.gs.washington.edu/EVS/>); (2) dbSNP (build 132); (3) St Jude/Washington University Paediatric Cancer Genome Project (PCGP); and (4) germline variants present in five or more TARGET CGI WGS cases in each cohort. Second, a variant was removed unless it met the following criteria: (1) at least three reads supported the mutant allele in the tumour; (2) the mutant read count in the tumour was significantly higher than normal ($P < 0.01$ by two-sided Fisher's exact test); and (3) the normal MAF was below 0.05. Finally, a BLAT search⁴⁰ was run on the mutant allele with 20-bp flanking to verify unique mapping.

A 'rescue' pipeline was implemented to avoid over-filtering, by using the customized AnnoVar annotation and pathogenicity identification tool Medal_Ceremony²³ (M.N.E. *et al.*, unpublished). Pathogenic variants were rescued and further curated with ProteinPaint²⁸.

This filtering has reduced the original 51 million SNVs and 38 million indels from the CGI files to a set of 711,490 SNVs and 57,700 indels. Of these, 9,397 SNVs and 1,000 indels are in protein coding regions. A comparison with gnomAD database (version r2.0.1; <http://gnomad.broadinstitute.org/>) indicated that 1.1% of our detected SNVs overlap with SNPs with population frequency greater than 0.1%. Verification of somatic point mutations after filtering is presented in Supplementary Note 1.

Filtering of structural variation. CGI structural variants were filtered to remove germline rearrangements, including those found in the Database of Genomic Variants, dbSNP, PCGP, recurrent germline rearrangements from CGI Mutation Annotation Format files, low-confidence somatic calls (>90% reference similarity to the assembled sequence) and those with both structural variant breakpoints falling into gap regions (hg19). Each structural variant was required to have an assembled contig length of at least 10 bp on each breakpoint. CNAs in each tumour were integrated into the structural variant analysis by matching breakpoints within a 5-kb window to rescue rearrangements with CNA support by manual curation. A comparison of CGI structural variants with the known oncogenic re-arrangement in AML and B-ALL is presented in Supplementary Note 2.

Copy number alterations. We adapted the CONSERING algorithm⁴¹ to detect CNAs from CGI WGS data. In brief, germline single nucleotide polymorphisms (SNPs) reported by CGI in Mutation Annotation Format files were extracted, and paralogous variants identified from 625 germline WGS cases generated by PCGP were removed. A coverage profile was constructed using the mean of SNP read counts within a sliding window of 100 bp, and the differences between tumour and normal samples were used as inputs for CONSERING. To detect LOH, we used SNPs with variant allele fraction (VAF) in normal sample within an interval

of (0.4, 0.6) and $>15\times$ coverage in tumour and normal samples. Allelic imbalance (AI; $|\text{Tumour_VAF}-0.5|$) was used to detect LOH. Regions with concomitant copy number changes ($|\log \text{ratio}| > 0.2$) or LOH (AI > 0.1) were subjected to manual review. Finally, regions less than 2 Mb were considered focal and included in the GRIN¹² analysis to determine the significance of the somatic alterations. A comparison of our CNA detection with clinical information is provided in Supplementary Note 3.

For osteosarcomas, manual reviews of candidate genes affected by CNA were prioritized for the following three groups owing to the high number of rearrangements caused by chromothripsis in this histotype⁴²: (1) gene expression change matched the CNA status; (2) genes with recurrent loss and gain; and (3) published osteosarcoma driver genes⁴². This resulted in the discovery of 13 focal CNAs affecting *CCNE1*, *CDKN2A*, *RB1*, *PTEN*, *TUSC7*, and *YAP1* in addition to *TP53*. **WES data analysis.** Of the 1,131 tumour-normal WES pairs, all but 23 osteosarcoma pairs exhibited the expected binomial distribution of B-allele fraction for germline SNPs. The 23 outlier samples were therefore used neither for the discovery of driver genes nor for calculating mutation rate in coding regions (Fig. 1b). They were included only for determining driver mutation prevalence.

Somatic SNVs and indels were detected by the Bambino⁴³ program, followed by postprocessing and manual curation as previously described^{44,45}. To address 8-oxo-G artefacts³⁵, we implemented the D-ToxoG filtering algorithm⁴⁶.

Somatic mutation rate. The median mutation rate of 651 CGI WGS samples (Fig. 1a) was calculated from tier3 non-coding SNVs⁴⁷. This analysis did not include the T-ALL cohort as only three T-ALLs were analysed by the CGI platform. Mutations in coding regions were based on coding SNVs from 1,639 samples analysed by WGS or WES (Fig. 1b). Among these, 120 samples were analysed by both WGS and WES, and the union of coding SNVs from WGS and WES were used. Twenty-three osteosarcoma WES samples were excluded from coding mutation analysis owing to quality issues described in 'WES data analysis'. For osteosarcomas, the mutation rate in coding regions (0.53 per Mb) is lower than in non-coding regions (0.79 per Mb). Nineteen osteosarcoma samples were analysed by both CGI and WES. For these samples, the mutation rate in coding regions derived from either CGI or WES was 0.54 per Mb while the mutation rate in the non-coding regions was 0.79 per Mb, indicating a potential contribution of kataegis⁴² in the elevated mutation rate in non-coding regions. Within each histotype, hypermutators were defined as having mutation rates 3 s.d. above the mean (trimming 5% outliers).

Mutational signature analysis. Mutational catalogues were generated for each sample by using a 96-bin classification (Supplementary Table 1b). These were examined for all samples with our previously established methodology⁴⁸ to decipher mutational signatures and to quantify their activities in individual samples. The correlation between age of diagnosis and mutational signature activities was computed by using robust regression⁴⁹. We also compared the cosine similarity between original and reconstructed samples and found that samples with more than 100 mutations had cosine similarities greater than 0.85, whereas samples with less than 100 mutations mostly (93.5%) had cosine similarities less than 0.85.

To calculate the average MAF values for each signature (Fig. 1e), each of the 96 mutation types was assigned to the signature with the highest probability (the same result was obtained if we required the highest probability to be higher than the second (by $\Delta = 0.05, 0.1$, and 0.2 ; data not shown). This assignment was also used for Extended Data Fig. 2e–i.

The two novel signatures, T-10 and T-11, were enriched in low MAF mutations. T-11 was the only signature that was significantly correlated ($r^2 = 0.9$) with the presence of multi-nucleotide variations composed of co-occurring SNVs separated by 3 or 4 bp which were not verified by Illumina WES. Therefore, it is likely to be associated with platform-specific sequencing artefacts.

For the eight B-ALL cases identified with mutation signatures of UV-light exposure, only 0.96% of the somatic SNVs overlap with SNPs that have population allele frequencies (AFs) $> 0.1\%$ in the gnomAD database (version r2.0.1; <http://gnomad.broadinstitute.org/>). The overlap is only 0.22% if using AF $> 1\%$. The overlap rate is comparable to the 1.1% observed for non-UV somatic SNVs across the entire cohort (0.27% match if using AF $> 1\%$).

For each of these eight B-ALL cases, UV- and non-UV-mutations were stratified according to the ploidy of their genomic locations (Extended Data Fig. 2e–g; cluster centres estimated using R package mclust). Inter-mutational distances were plotted for comparison of genomic distribution of UV- versus non-UV mutations. Chromosomal ploidy and tumour purity were obtained from TARGET clinical files and prior publications⁵⁰. By adjusting for ploidy and corresponding tumour purity, we calculated expected MAFs for clonal mutations as follows: denoting the tumour purity as π , the expected MAF for clonal mutations was $\pi/(2 - \pi)$ in the 1-copy loss region, $\pi/2$ in the diploid region, and $\pi/(2 + \pi)$ in the 1-copy gain (wild-type allele) region.

Age-specific incidence rates for childhood ALL reported by the Surveillance, Epidemiology, and End Results (SEER) program show that the rate of incidence

in African American children is half of that in white children (Extended Data Fig. 2h). While none of our eight patients is African American according to clinical information and genomic imputation, we were not able to test the significance of this observation as 6.6% of the children enrolled in the COG ALL trial are African American.

Chromothripsis analysis. To detect chromothripsis, we first assessed whether the distribution of structural variant breakpoints in each tumour departed from the null hypothesis of random distribution using Bartlett's goodness-of-fit test⁴². The distribution of structural variant types (deletion, tandem duplication, head-to-head and tail-to-tail rearrangements) was also evaluated using a goodness-of-fit test for chromosomes with a minimum of five structural variants. Chromosomes with $P < 0.01$ for Bartlett's test and with $P > 0.01$ for the structural variant type test were further reviewed for oscillation between restricted CNA states.

Discovery of candidate driver genes. For the 654 CGI samples, we ran GRIN¹² analysis with all somatic variants (structural variants, CNAs, SNVs and indels) for both individual histotypes and a combined pan-cancer cohort. Similarly, we combined coding SNVs and indels identified in both WGS and WES for MutSigCV¹³. Putative genes with $Q < 0.01$ by GRIN or MutSigCV were subjected to additional curation to determine their driver status. Only one candidate gene was included in this analysis for somatic alterations affecting multiple genes such as fusion pairs (Supplementary Note 4).

We discovered 142 candidate driver genes by this approach (Supplementary Table 2). Of these, 133 were significant by GRIN analysis (87 genes common to both GRIN and MutSigCV) and nine were significant only by MutSigCV.

HotNet2 analysis. We applied HotNet2³ to somatic mutations using interaction data obtained from the HINT, HG2014, and KEGG databases. We reviewed all predicted sub-networks and identified the cohesin complex with three additional genes (*STAG1*, *PDS5A* and *PDS5B*; Extended Data Fig. 4e, f).

Pathway analysis. Biological pathways for candidate driver genes were assigned using public pathway databases (KEGG and version 2.0 of the NCI RAS Pathway, https://www.cancer.gov/PublishedContent/Images/images/nci/organization/ras/blog/ras-pathway-v2_v60096472.jpg), literature reviews, and biological networks produced by HotNet2. For each pathway in each histotype, a tumour was counted if any genes of that pathway were mutated. The percentage of variants in genes unique to paediatric cancers was calculated by excluding genes reported in the three TCGA pan-cancer studies^{1,2,4}.

Mutual exclusivity and co-occurrence of mutations. We tested mutual exclusivity and co-occurrence of mutations for the 142 driver genes. For each histotype, we performed this analysis in two separate sample sets: (1) samples with WGS (T-ALL with WES and SNP6), and (2) WGS and WES together (only SNVs and indels considered to avoid detection bias due to platform differences for CNVs and structural variants). For a gene pair A and B (mutated in five or more samples), we performed two-sided Fisher's exact test according to their mutation status. The R package *qvalue*⁵¹ was used to control for multiple testing. Although the co-occurrence test is well-powered for most gene pairs, we recognize that the mutual exclusivity test is not powered for most gene pairs, and pairs with $P < 0.05$ were reported even if $Q > 0.05$ (Supplementary Table 3).

Saturation analysis. To study the effect of sample size on detecting driver genes, we performed down-sampling analysis in the pan-cancer cohort and in each histotype², for GRIN and MutSigCV separately. For each combination, we repeated the statistical analysis for a series of subsets of cases from 1 to the total number of samples. The number of genes (of the 142 driver genes) with false discovery rate less than 0.01 were counted for the corresponding subset. Analysis for individual histotypes was limited to those with at least 200 samples (osteosarcomas and Wilms tumours excluded).

Somatic variant pathogenicity analysis. We implemented a somatic mutation classifier Medal_Ceremony²³ (M.N.E. *et al.*, unpublished) to identify additional driver variants in genes that did not pass the statistical testing. Pathogenic variants include (1) hotspot SNVs and indel mutations for known cancer genes in any cancer type; (2) pathogenic mutations in ClinVar; (3) truncation mutations in known tumour suppressor genes that were expressed in the cancer histotype; and (4) known recurrent gene fusions, focal deletions, truncations, and amplifications that affect key pathways of any cancer type and that were simultaneously corroborated by an aberrant expression profile. We identified 184 variants in another 82 genes (Supplementary Table 4). *BRAF* was the most frequently mutated, with nine variants.

We also reviewed novel hotspot mutations detected in three or more samples. After removing low-confidence mutations and those without expression, one hotspot was found (*MAP3K4* G1366R, $n = 4$). Recurrent internal tandem duplication (ITD) was also reviewed for evidence in both DNA and RNA, yielding the discovery of *UBTF*-ITD in AML.

Tumour purity assessment. We used regions with copy number loss or copy neutral LOH as well as SNVs (coding and noncoding) from diploid regions to

estimate tumour purity. For regions with LOH, a previously described method was used⁴². For SNVs, an unsupervised clustering analysis was performed with the R package *mclust*. Tumour purity was defined to be two times the highest cluster centre that was < 0.5 . The maximal CNA and SNV purity was used.

We compared our estimates with blast counts for 197 AML and 9 B-ALL samples. Of the 135 tumours with blast count $> 70\%$ (value 'many' in clinical file was mapped as $> 70\%$), we identified 127 (94%) with purities $> 70\%$ (seven of the other eight tumours had purities $> 50\%$). An additional 40 tumours were estimated with purities $> 70\%$, although their blast count was below 70%. Thirty-one tumours were classified as low purity ($< 70\%$) by both our analysis and blast count.

KRAS isoform analysis. We investigated alternative splicing in *KRAS* (Extended Data Fig. 6), as differential oncogenic activity of mutant alleles expressed in *KRAS* 4a or 4b isoforms has been reported previously⁵². We detected splice junction reads connecting exon 3 to one of the two novel acceptor sites in the last intron (53 bp apart). This aberrant splicing is predicted to create two novel isoforms, each incorporating one of the two novel exons (40 bp and 93 bp, respectively) located 2.2 kb downstream of exon 4A (Extended Data Fig. 6b). These novel isoforms would form truncated *KRAS* proteins (154/150 amino acid), each retaining the GTPase domain but losing the hypervariable region that is critical for targeting *KRAS* to the plasma membrane⁵³.

One of the two novel isoforms (novel isoform 2) was detected in myeloid cells from three healthy donors (data not shown). Protein products of *KRAS* isoforms in AML cells were analysed by western blot (Supplementary Notes 5, 6).

RNA-seq data analysis. RNA-seq data were mapped with StrongArm²³, and rearrangements identified with CICERO²³, followed by manual review. We performed RNA-seq clustering to confirm the tissue of origin and analysed immune infiltration using ESTIMATE⁵⁴ and CIBERSORT⁵⁵ (Extended Data Fig. 7, Supplementary Notes 7, 8).

Allele-specific expression (ASE) in RNA-seq. CGI and WES allele counts were combined whenever possible. Point mutations were required to have DNA and RNA coverage $\geq 20 \times$. Variants with $|RNA_MAF - DNA_MAF| > 0.2$ and a false discovery rate of < 0.01 (calculated with R package *qvalue*⁵¹ on two-sided Fisher's exact test P) were considered to show ASE. Within-sample analysis was performed to distinguish ASE from potential artefacts caused by normal-in-tumour contamination (Extended Data Fig. 8d, Supplementary Note 9, Supplementary Table 5).

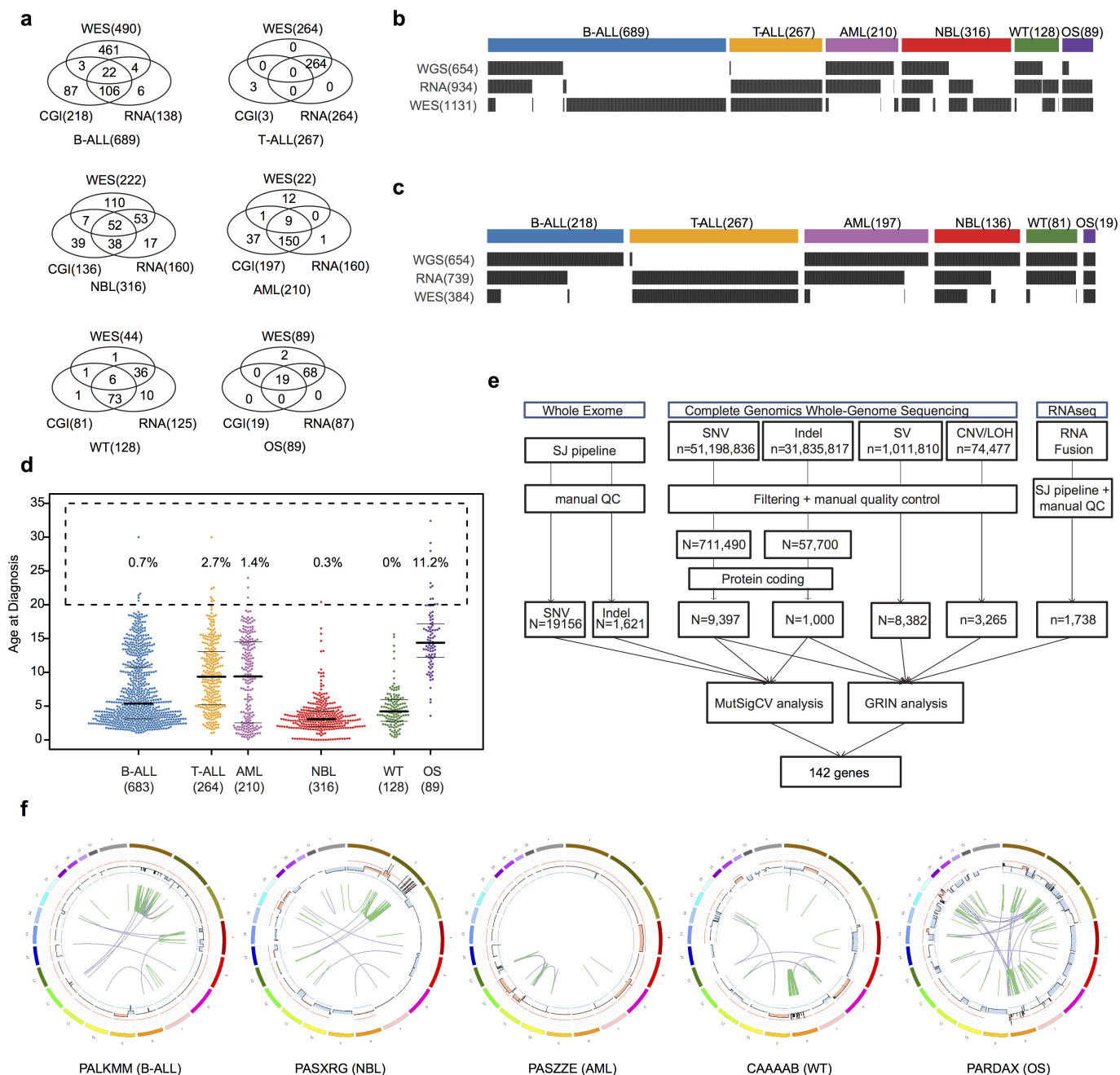
Single-cell targeted re-sequencing. One cryopreserved vial each from patients PAPWIU and PABLDZ was thawed using the ThawSTAR system (MedCision) and then diluted in RPMI supplemented with 1% BSA. The cells were then washed five times with C1 DNA-seq wash buffer according to the manufacturer's instructions (Fluidigm), counted and viability estimated using the LUNA-FL system (Logos Biosystems), then diluted to 300 cells per μ l and loaded in a small C1 DNA-seq chip according to the manufacturer's instructions (Fluidigm, except the suspension buffer to cell ratio was changed from 4:6 to 6:4). The cells also underwent an on-chip LIVE/DEAD viability stain (Thermo Fisher). Each capture site was imaged using a Leica inverted microscope and phase contrast images, as well as fluorescent images with GFP and Y3 filters, were acquired to determine the number of cells captured and the viability of each. The cells then underwent lysis, neutralization, and MDA WGA according to the manufacturer's instructions (Fluidigm) using the GenomePhiv2 MDA kit (GE Life Sciences). One C1 chip was run per patient. Selected variants and germline SNPs then underwent microfluidic PCR-based targeted resequencing in the bulk sample or genomes amplified from the single cells using the Access Array System as previously described⁵⁶. Target-specific assays were designed using primer3plus (<https://probes.pw.usda.gov/batchprimer3/>) and employed oligos purchased from Integrated DNA Technologies; multiplexing was performed according to guidelines in the Access Array manual (Fluidigm). All samples were loaded with the Access Array loader and underwent PCR cycling in an FC1 system, followed by sample-specific barcoding using standard PCR, all according to the manufacturer's instructions (Fluidigm). Amplicons were run on the MiSeq using v2 chemistry with 2×150 -bp paired-end reads (Illumina), using custom sequencing primers, according to the Access Array manual (Fluidigm).

Single-cell sequencing data analysis. Mapped BAM files for each of the 96 single-cell assays were genotyped for all designed markers. Assays with two captured cells (6 assays for both cases) or assays with fewer than 50% of designed markers with coverage $10 \times$ or greater, were dropped, resulting in 48 assays for case PAPWIU (Supplementary Table 6) and 64 assays for case PAPEWB (Supplementary Table 7). The assays were called tumour cells if they had one or more somatic markers with MAFs greater than 0.05. Germline markers with MAFs greater than 0.05 were called positive. The R package *phatmap* was used to visualize the single-cell data using hierarchical clustering with 'binary' distance and 'complete' agglomeration method.

Code availability. Custom codes are available from the authors upon request.

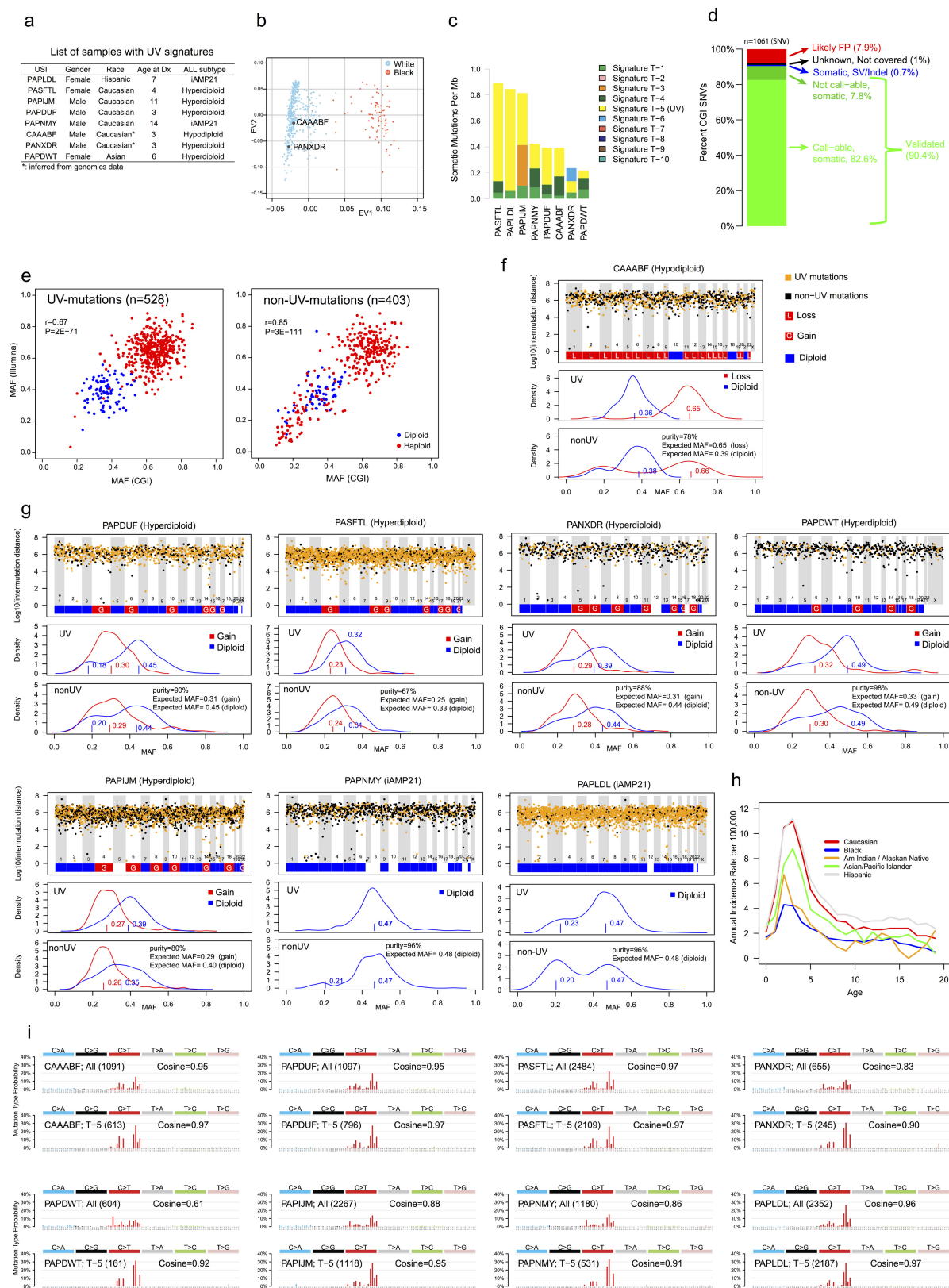
Data availability. The somatic variants used for this study are available at the National Cancer Institute TARGET Data Matrix (<https://ocg.cancer.gov/programs/target/data-matrix>) and our ProteinPaint²⁸ portal (<https://pecan.stjude.org/proteinpaint/study/pan-target>), which also hosts variant data generated by Gröbner *et al.*²⁹ (<https://pecan.stjude.org/proteinpaint/study/dkzf-ppc>).

30. Mirabello, L., Troisi, R. J. & Savage, S. A. International osteosarcoma incidence patterns in children and adolescents, middle ages and elderly persons. *Int. J. Cancer* **125**, 229–234 (2009).
31. Behjati, S. *et al.* Recurrent mutation of IGF signalling genes and distinct patterns of genomic rearrangement in osteosarcoma. *Nat. Commun.* **8**, 15936 (2017).
32. Mullighan, C. G. *et al.* Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. *N. Engl. J. Med.* **360**, 470–480 (2009).
33. Ma, X. *et al.* Rise and fall of subclones from diagnosis to relapse in pediatric B-acute lymphoblastic leukaemia. *Nat. Commun.* **6**, 6604 (2015).
34. Bolouri, H. *et al.* The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. *Nat. Med.* **24**, 103–112 (2018).
35. Pugh, T. J. *et al.* The genetic landscape of high-risk neuroblastoma. *Nat. Genet.* **45**, 279–284 (2013).
36. Gadd, S. *et al.* A Children's Oncology Group and TARGET initiative exploring the genetic landscape of Wilms tumor. *Nat. Genet.* **49**, 1487–1494 (2017).
37. Drmanac, R. *et al.* Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* **327**, 78–81 (2010).
38. Carnevali, P. *et al.* Computational techniques for human genome resequencing using mated gapped reads. *J. Comput. Biol.* **19**, 279–292 (2012).
39. Lee, W. *et al.* The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature* **465**, 473–477 (2010).
40. Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
41. Chen, X. *et al.* CONSERTING: integrating copy-number analysis with structural-variation detection. *Nat. Methods* **12**, 527–530 (2015).
42. Chen, X. *et al.* Recurrent somatic structural variations contribute to tumorigenesis in pediatric osteosarcoma. *Cell Rep.* **7**, 104–112 (2014).
43. Edmonson, M. N. *et al.* Bambino: a variant detector and alignment viewer for next-generation sequencing data in the SAM/BAM format. *Bioinformatics* **27**, 865–866 (2011).
44. Zhang, J. *et al.* The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature* **481**, 157–163 (2012).
45. Zhang, J. *et al.* A novel retinoblastoma therapy from genomic and epigenetic analyses. *Nature* **481**, 329–334 (2012).
46. Costello, M. *et al.* Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic Acids Res.* **41**, e67 (2013).
47. Mardis, E. R. *et al.* Recurring mutations found by sequencing an acute myeloid leukemia genome. *N. Engl. J. Med.* **361**, 1058–1066 (2009).
48. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep.* **3**, 246–259 (2013).
49. Alexandrov, L. B. *et al.* Clock-like mutational processes in human somatic cells. *Nat. Genet.* **47**, 1402–1407 (2015).
50. Holmfeldt, L. *et al.* The genomic landscape of hypodiploid acute lymphoblastic leukemia. *Nat. Genet.* **45**, 242–252 (2013).
51. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA* **100**, 9440–9445 (2003).
52. To, M. D. *et al.* Kras regulatory elements and exon 4A determine mutation specificity in lung cancer. *Nat. Genet.* **40**, 1240–1244 (2008).
53. Eisenberg, S. & Henis, Y. I. Interactions of Ras proteins with the plasma membrane and their roles in signaling. *Cell. Signal.* **20**, 31–39 (2008).
54. Yoshihara, K. *et al.* Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612 (2013).
55. Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).
56. Gawad, C., Koh, W. & Quake, S. R. Dissecting the clonal origins of childhood acute lymphoblastic leukemia by single-cell genomics. *Proc. Natl Acad. Sci. USA* **111**, 17947–17952 (2014).



Extended Data Figure 1 | Cohort description and workflow. **a**, Venn diagram of samples analysed by whole-exome (WES), whole genome (CGI) and whole transcriptome (RNA-seq) sequencing in this cohort. **b**, **c**, Sample-level sequencing status of the entire cohort (**b**) and those with WGS data (**c**, SNP6 for T-ALL). **d**, Age distribution for each histotype. Median, first and third quartiles are indicated by horizontal bars. Sample sizes are indicated in parentheses. Percentage of cases with age over 20 years are indicated. **e**, Analytical workflow. The tumour/normal BAM files of WES data were analysed by our in-house pipeline followed by manual quality control. The mutation annotation format files generated by CGI were downloaded from TARGET Data Matrix (see Methods) and analysed by a pipeline developed for this dataset, including SNVs,

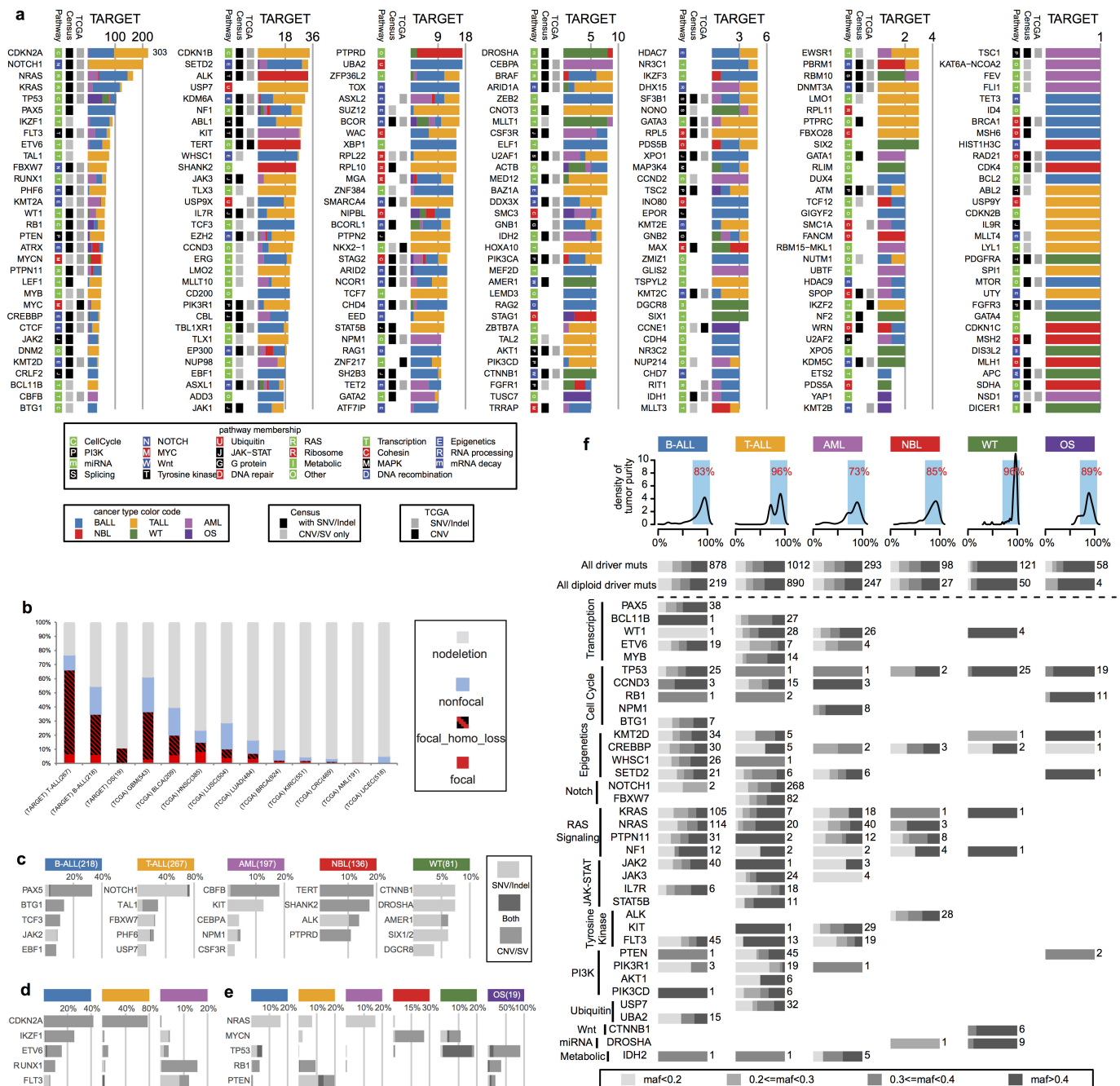
indels and structural variants. CAN and LOH were analysed using read counts of germline SNPs in the mutation annotation format files. Manual quality control was also performed. For RNA-seq data, the FASTQ files were re-mapped and fusions and ITDs were analysed with CICERO. The resultant mutations were analysed by GRIN (SNVs, indels, CNAs, structural variants and fusions) and MutSigCV (SNVs and indels) to discover 142 recurrently mutated genes. **f**, One representative sample with chromothripsis for each histotype. CNAs are shown in the inner circle, orange indicates copy gain and blue indicates copy loss. Intra- and inter-chromosomal rearrangements are shown as green and purple curves, respectively.



Extended Data Figure 2 | See next page for caption.

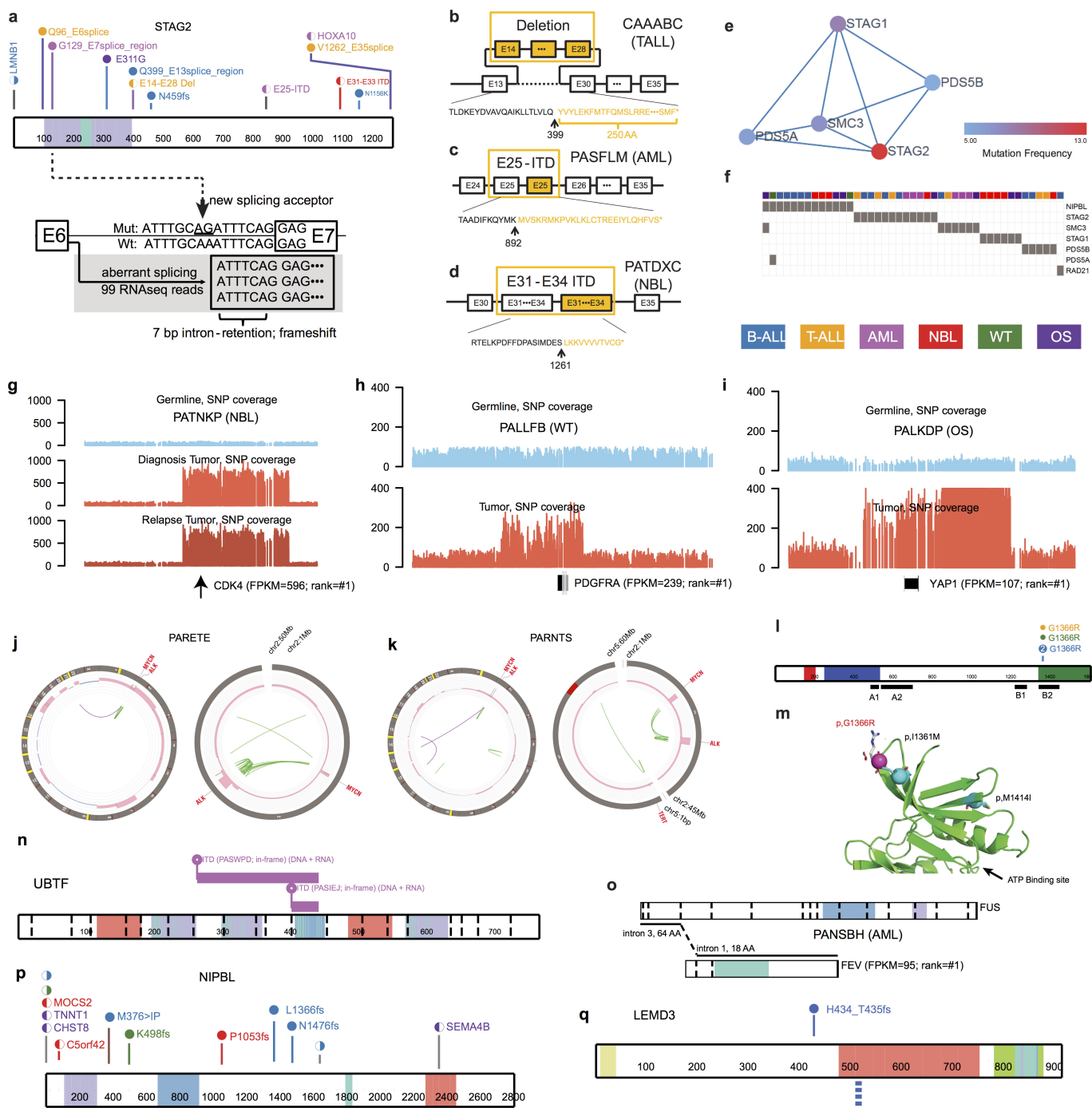
Extended Data Figure 2 | Eight B-ALL samples with signatures of UV exposure. **a**, List of samples with UV signatures detected. **b**, Inference of ethnicity for cases CAAABF and PANXDR from 654 TARGET CGI samples by principal component analysis (Supplementary Note 10). **c**, Total spectrum of mutational signatures of the eight UV mutation samples. **d**, SNVs of case CAAABF have a cross-validation rate of 90.4% with Illumina WGS data. **e**, High concordance of MAF values of SNVs derived from CGI and Illumina WGS, categorized by UV and non-UV mutations. Listed are Pearson's correlation coefficient (r) and P value derived from linear regression. Numbers of SNVs are indicated in parentheses. **f**, Inter-chromosomal distance and density plots for UV and non-UV mutations in case CAAABF. Top, inter-mutational distance (\log_{10} scale) of UV (orange dots) and non-UV (black dots) mutations. Chromosomal level gain and loss statuses are indicated. The results indicate uniform distribution of mutations with or without UV signature

across the genome. Middle and bottom panels show density plots of UV- and non-UV-mutations, respectively, categorized by chromosomal loss (red) and diploid (blue) status in corresponding tumour samples. Estimated cluster centres are indicated by corresponding colours. The expected MAFs for clonal mutations at given purity and chromosomal ploidy status of corresponding tumour are listed in the bottom panel. The density plots show that mutations with UV signatures are clonal after adjusting for ploidy. **g**, Inter-chromosomal distance and density plots for the other seven cases (key shown in **f**). **h**, ALL incidence by ethnicity obtained from the most recent registry (1973–2014) of SEER Program research data (Supplementary Note 11). **i**, Mutation spectrum for all SNVs (All) and for UV SNVs (T-5) for each of eight cases. Total number of SNVs and cosine similarity with COSMIC signature-7 are indicated in each panel.



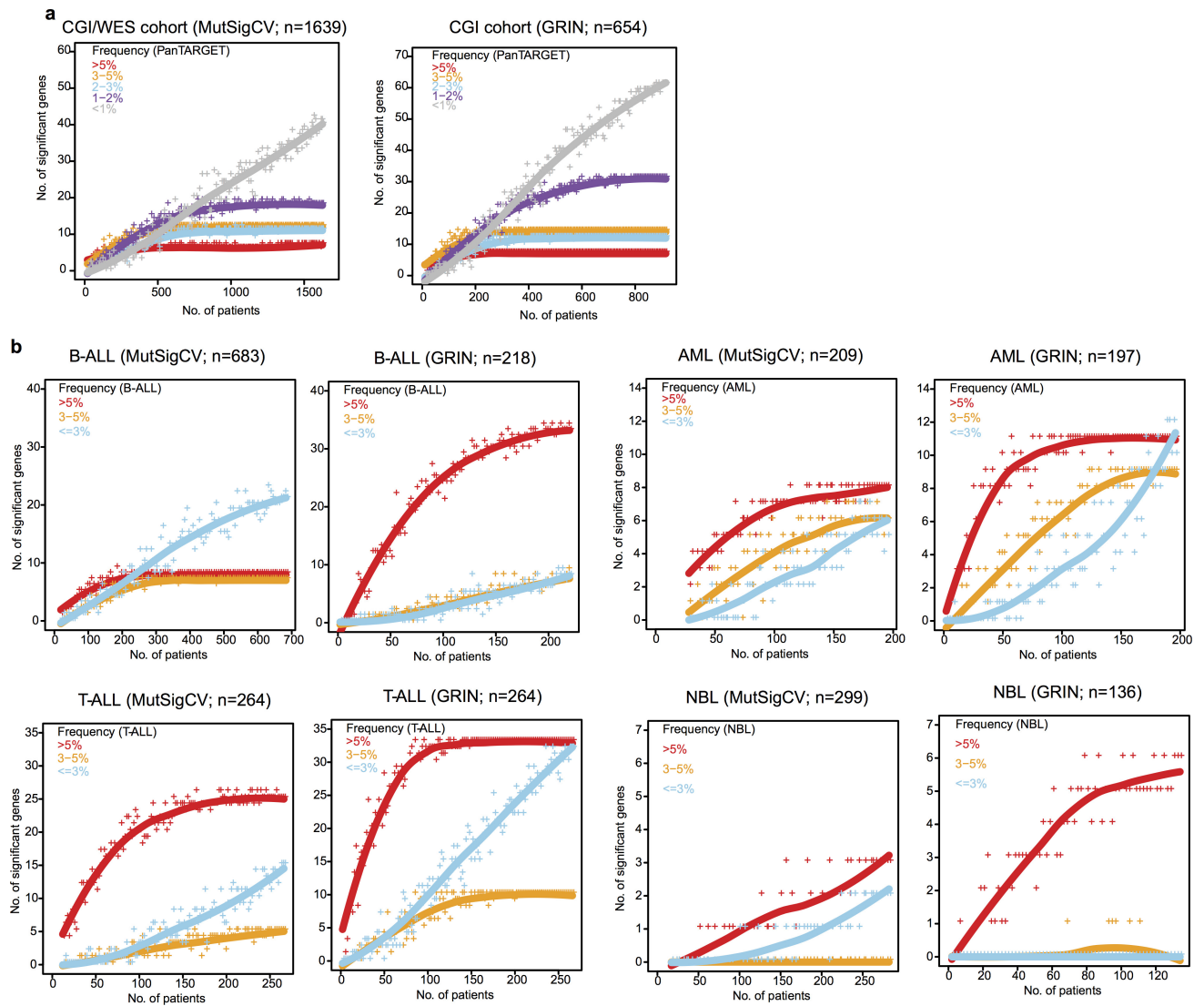
Extended Data Figure 3 | Driver mutation landscape in paediatric cancers. a, The number of samples mutated in each histotype is shown with colours coded as in Fig. 2. The presence of each gene in the Cancer Gene Census (Census) and prior pan-cancer studies of The Cancer Genome Atlas (TCGA) project are indicated. Pathway membership is also labelled for each gene. Somatic alterations in T-ALL were based on coding SNVs and indels from WES and CNAs from SNP array. **b**, Percentage of samples with focal (≤ 2 Mb) and non-focal (> 2 Mb) deletions in *CDKN2A*. In the focal deletion category, samples with a second hit (either a second CNA or a copy neutral LOH) were categorized as 'focal_homo_loss'. For B-ALL, 27 of 218 (12%) non-focal samples had arm-level (such as hyperdiploid or hypodiploid B-ALL) CNAs on chromosome 9. Nine of 218 (4%) B-ALL cases had homozygous *CDKN2A* deletions with sizes from 2.1 Mb to 7.2 Mb and were counted as non-focal. TCGA data (no ALL data

available) were downloaded in December 2015. The number of samples is indicated for each histotype. **c**, Top five genes mutated exclusively in each histotype. **d**, Top five genes mutated in leukaemias. **e**, Top five genes mutated in both leukaemias and solid tumours. **f**, MAF distribution of point mutations in driver genes. Top, density plot of tumour purity for each histotype. Percentages of samples with tumour purity $> 70\%$ are indicated. Bottom, MAF distribution of point mutations in driver genes. Aggregated distribution for all driver genes is shown at the top ('All driver muts'), as well as all driver genes in diploid regions (for CGI data, CNA $|\text{seg.mean}| < 0.2$, $|\log\text{Ratio}| < 0.2$, and LOH $\text{seg.mean} < 0.1$; for T-ALL SNP array data, CNA $|\text{seg.mean}| < 0.2$). For each biological process defined in Fig. 3, the MAF distribution is shown for the genes with the five highest mutation frequencies that are mutated in more than five samples. The number of mutations in each histotype is shown.



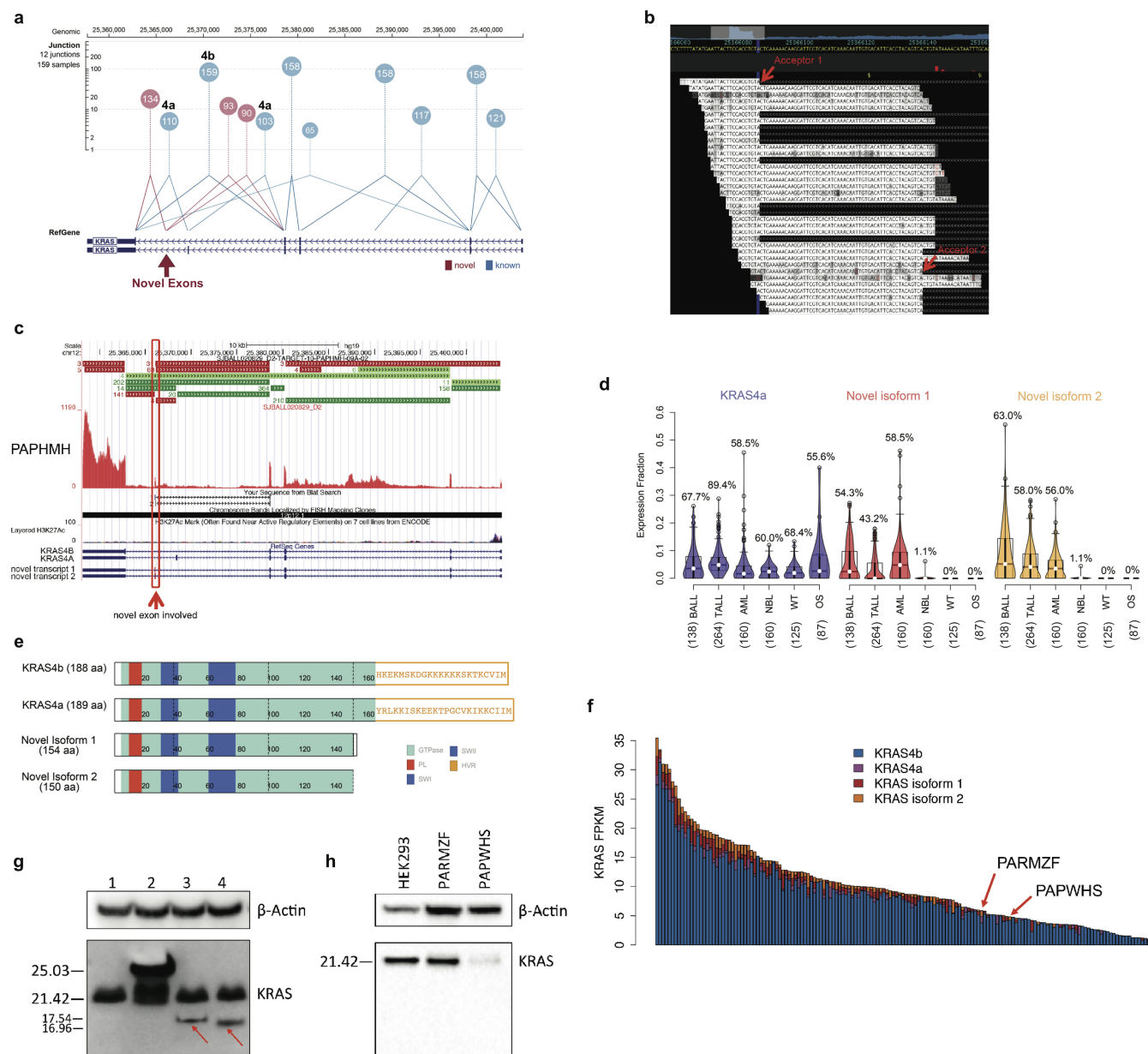
Extended Data Figure 4 | Example driver mutations. **a**, Diverse mutation types of *STAG2*. Variants are coloured by histotype as in Fig. 2. Circles and half-moons represent mutations and structural alterations, respectively. Bottom panel shows RNA-seq for an SNV at the -8 position of *STAG2* exon 7, which created a *de novo* splice site resulting in an out-of-frame transcript. **b-d**, Truncating mutations by deletion or ITD. **e**, Cohesin complex detected by HotNet2 analysis. **f**, Samples with mutations in cohesin complex. **g-k**, Selected examples of singleton oncogenic activation caused by high level amplifications including

CDK4 (**g**), *PDGFRA* (**h**), and *YAP1* (**i**) with FPKM and histotype-wise ranks indicated, as well as recurrent co-amplification of *MYCN-ALK* in two NBL samples (**j**, **k**). **l**, Recurrent *MAP3K4* mutation with structural model in N lobe (**m**). Location of the mutation p.G1366R is indicated by a magenta sphere and the alteration side chain is modelled as a stick. Known activating alterations (p.I1361M and p.M1415I) are shown as teal spheres. GADD45 binding (A1), kinase inhibitor (A2), and kinase domains (B1, B2) are indicated in **l**. **n**, ITD in *UBTF*. **o**, Fusion of *FEV*. **p**, **q**, Mutations in novel driver genes *NIPBL* and *LEMD3*.



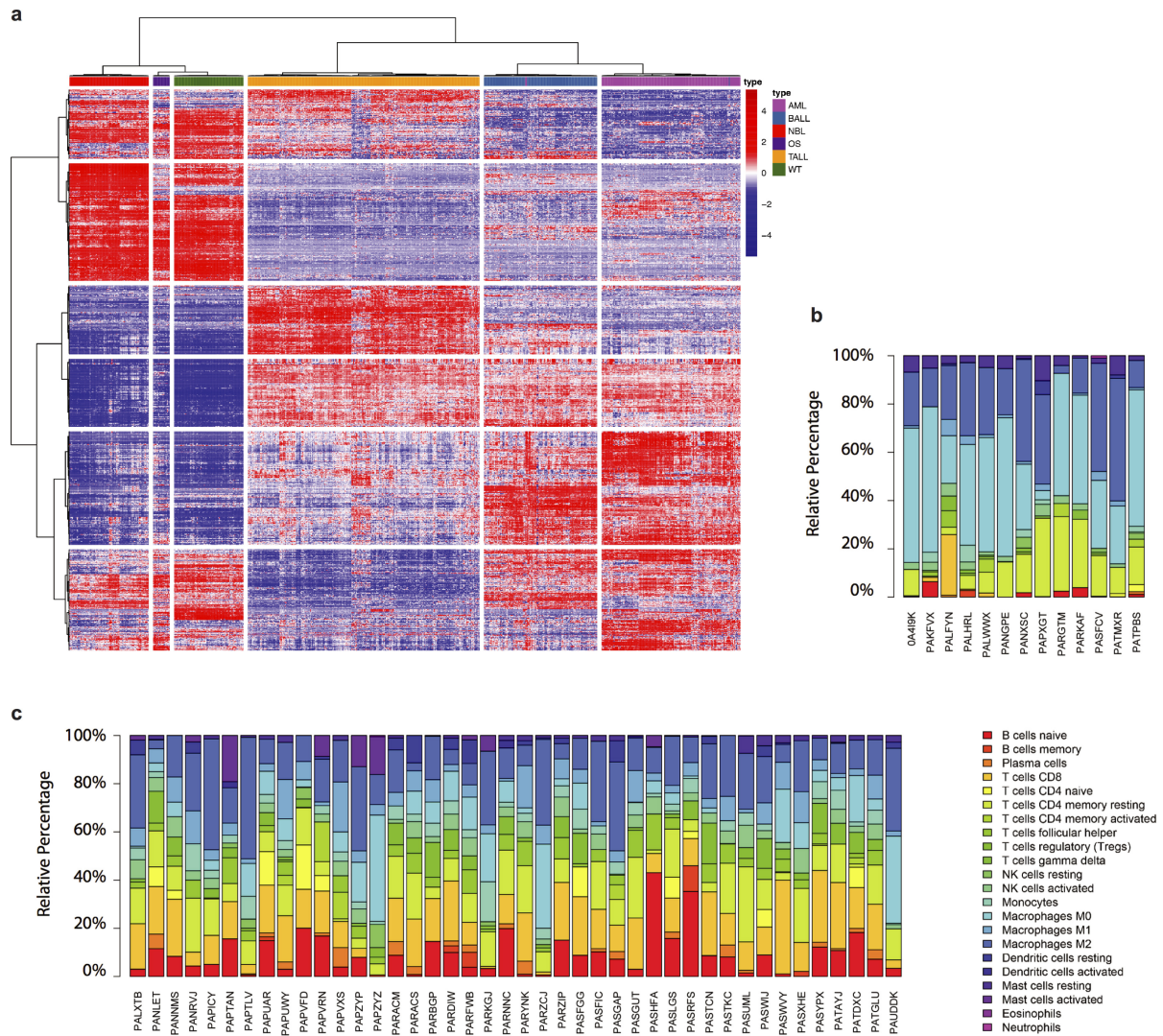
Extended Data Figure 5 | Down-sampling analysis of gene discovery. The analysis was performed on point mutations with MutSigCV and on SNVs, indels, structural variants, CNAs and fusion variants with GRIN (see Methods). The resulting candidate driver genes were categorized into five frequency bins indicated by different colours. Each point (+) represents a random subset of the pan-cancer cohort. The line is a

smoothed fit. **a**, Analysis performed on entire CGI/WES cohort with MutSigCV (left) and CGI cohort with GRIN (right). **b**, Analysis performed with MutSigCV and GRIN for each histotype. Candidate driver genes were assigned to three frequency bins (according to corresponding histotypes). Sample sizes are indicated in parentheses in each panel.



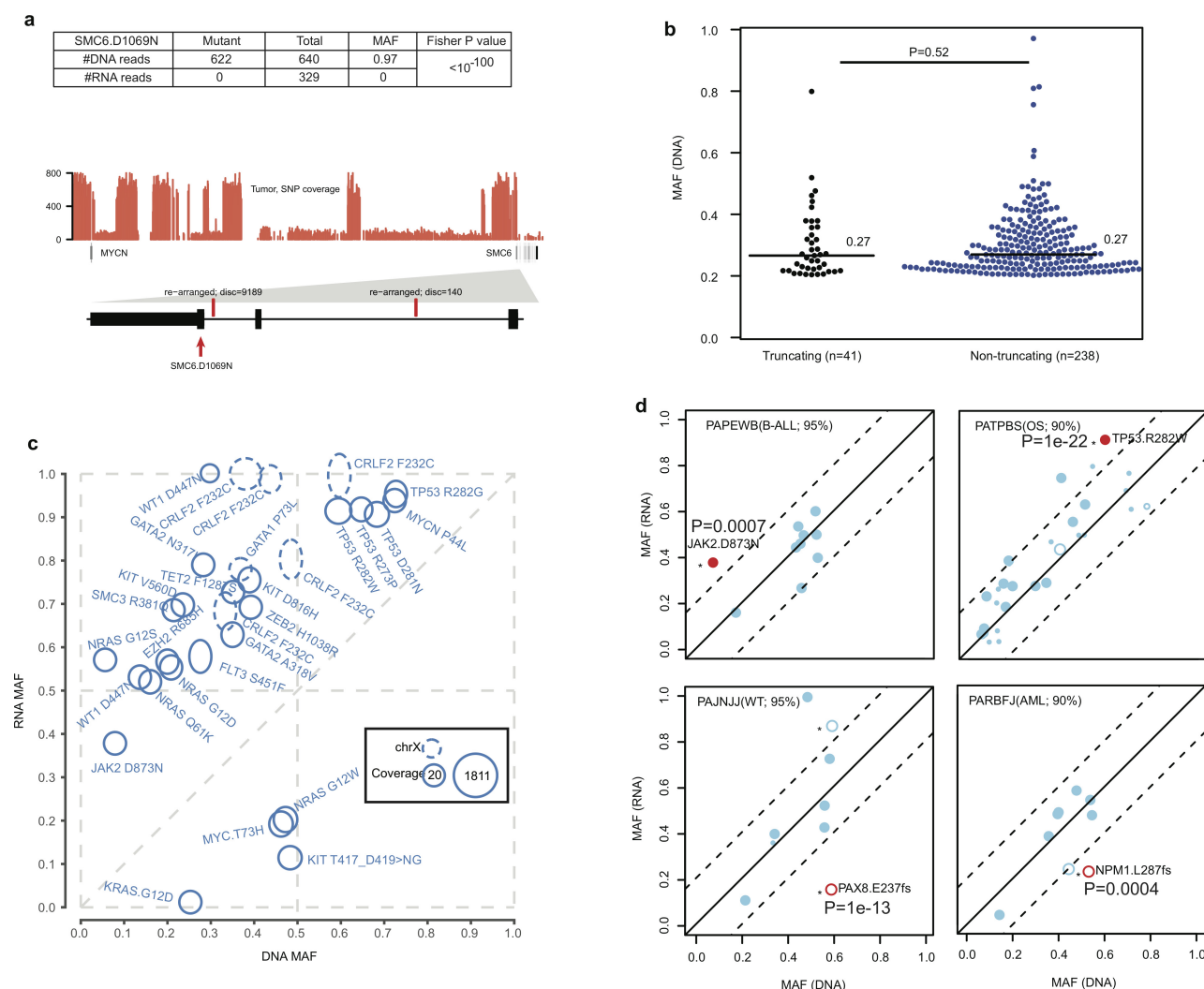
Extended Data Figure 6 | Expression of novel KRAS isoforms. **a**, KRAS RNA-seq reads spanning splice junctions in AML samples. Each junction is shown as a circle labelled by counts of detected samples, with lines connecting the splice sites. The circle's y-axis position represents the median supporting read count. Canonical junctions are coloured blue and novel junctions red. **b**, RNA-seq reads in the last intron of KRAS illustrate the two novel exons detected in a B-ALL sample (PAPMHM). Novel splicing acceptor sites are indicated by red arrows. **c**, Junction reads for KRAS in the same B-ALL sample. Canonical KRAS exons are shown as green horizontal bars while novel exons are shown in red (top) and the RNA-seq coverage at the KRAS gene locus is shown below. The two novel exons are indicated by red arrows. **d**, Expression of two novel isoforms with KRAS4a as a control. Percentage of samples expressing these isoforms

are indicated. Median, first and third quartiles are indicated by horizontal bars. Sample sizes are indicated in parentheses. **e**, Protein domains for KRAS4a, KRAS4b and two novel isoforms. **f**, KRAS expression (FPKM) in AML samples analysed in this study, categorized by the four isoforms. **g**, Western blot for KRAS in 293T cells. Cells were transfected with empty vector (lane 1), tagged wild-type KRAS (lane 2), novel isoform 1 (lane 3) and novel isoform 2 (lane 4). Protein products of the two novel KRAS isoforms are indicated by red arrows. **h**, Western blot for KRAS in two patient tumour samples (PARMZf and PAPWHS). Protein products of the two novel isoforms were not detected in these two samples. For **g** and **h**, the experiments were performed in duplicate and similar results were observed (see Supplementary Fig. 1 for gel source data).



Extended Data Figure 7 | Clustering analysis of tumour RNA-seq data and immune cell infiltration analysis. **a**, Clustering analysis was carried out for 739 primary tumours with RNA-seq data available. The top 1,000 most variable expressed genes were clustered using Ward's minimum variance method. Each disease is annotated as shown in the first row with

colour indicated in the key. **b**, **c**, Immune cell infiltration in osteosarcomas and NBL. Macrophage M0 and M2 were the dominant immune cell populations observed in osteosarcoma tumours (**b**). T and B cells, followed by macrophages, were the major immune cell types observed in NBL tumours (**c**).



Extended Data Figure 8 | Analysis of allele-specific expression.

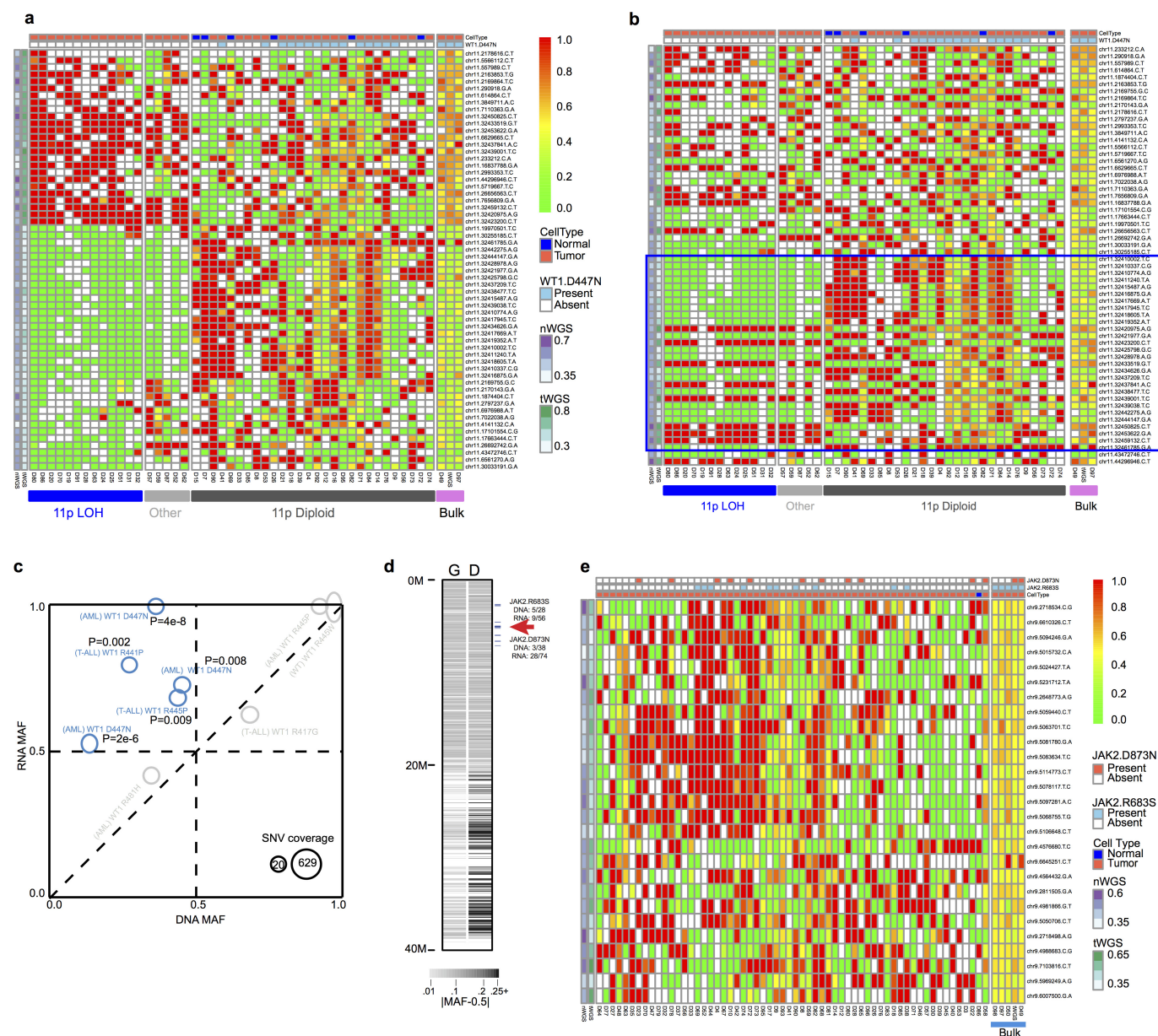
a, Mutant allele and total read count for SMC6 D1069N in DNA and RNA of NBL case PAPZYP. This is to illustrate variants with suppressed mutant allele expression despite high DNA MAF and a high level of gene expression in RNA-seq. *P* value was calculated using two-sided Fisher's exact test. DNA coverage of the *MYCN* and *SMC6* region indicating multiple segments with high amplification (estimated at 26 copies). Details of the last three exons (E26, E27 and E28) of *SMC6* are shown with DNA structural variants highlighted by vertical red bars. The mutation SMC6 D1069N is present in a region disrupted by structural variants, which dissociate the last three exons from the rest of *SMC6*. The high DNA MAF was therefore within a gene fragment that could not be transcribed and the expressed reference allele was from the intact gene.

b, Non-expressed truncating (black) and non-truncating (blue) mutations showed a similar ($P = 0.52$, two-sided Wilcoxon rank-sum test) median MAF (horizontal black lines). The number of SNVs in each category is shown in parentheses.

c, Hot spot mutations exhibited elevated mutant allele expression. Each mutation is shown as an oval positioned by its DNA

MAF (*x* axis) and RNA MAF (*y* axis). The read count in DNA and RNA is depicted by the radius along the *x*-axis and *y*-axis direction, respectively. Mutations on chromosome X are shown as dotted ovals. Read counts from CGI and WES were combined whenever possible.

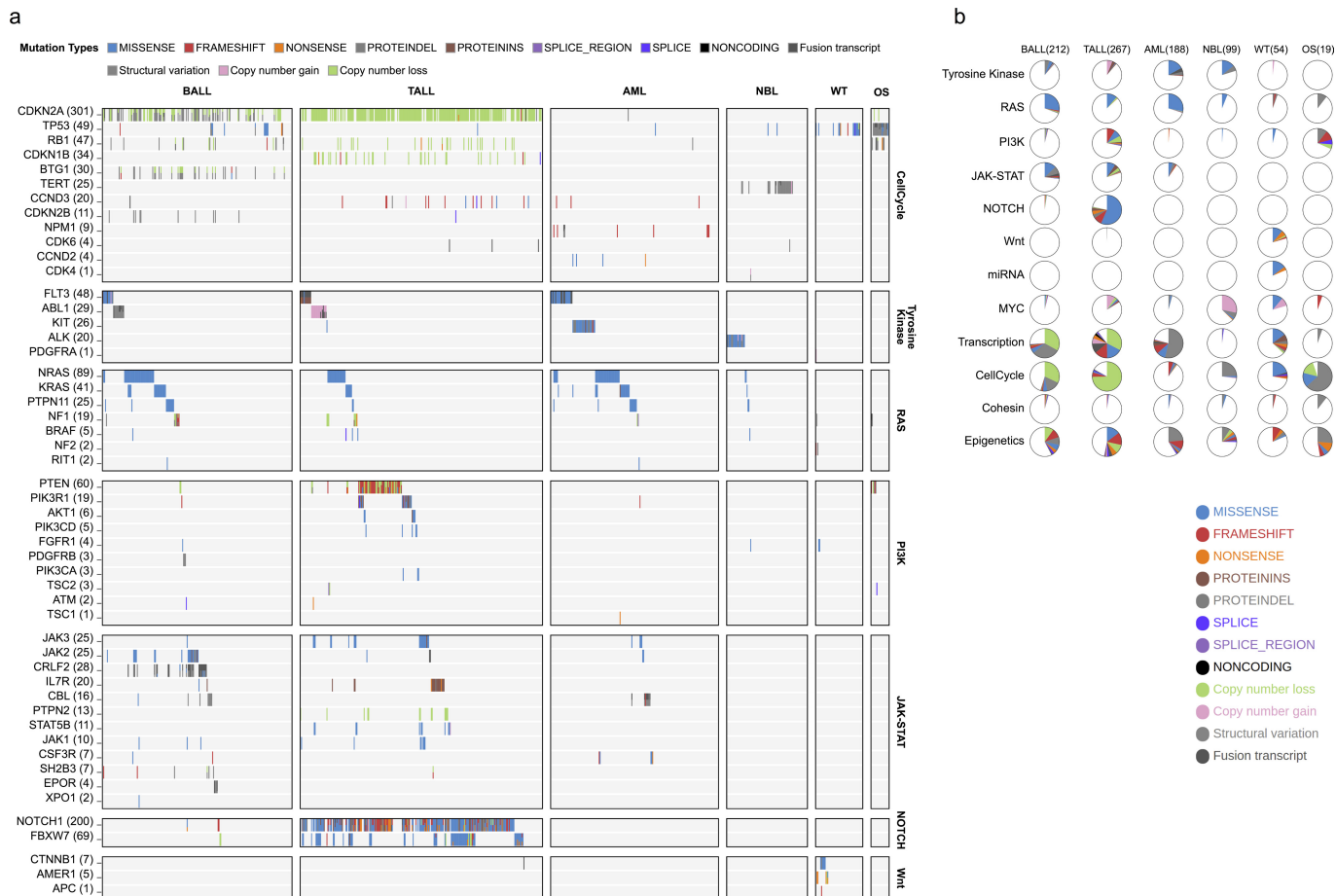
d, Within-sample analysis to evaluate the effect of normal cell contamination on ASE. Shown are two samples with hotspot SNVs (red dots in cases PAPEWB and PATPBS) and two samples with truncating mutations (red circles in cases PAJNJ and PARBFJ), which had a sufficient number of expressed coding mutations. The purity of each tumour is indicated. Dots represent SNVs and circles represent indels. Smaller symbols indicate the presence of CNA or LOH. An asterisk indicates a significant difference in MAFs between DNA (*x* axis) and RNA (*y* axis), which requires a minimum MAF difference of 0.2 (dashed lines) and a two-sided Fisher's exact test $P < 0.01$ (exact *P* values indicated in each panel). A dot in case PAJNJ with DNA MAF of 0.5 and RNA MAF of 1.0 is not significant owing to low coverage ($2\times$) in RNA-seq. In all four cases, within-sample concordance of DNA and RNA MAF for all except the ASE mutation suggest that normal cell contamination has a negligible effect on ASE.



Extended Data Figure 9 | Allele-specific expression of *WT1* and *JAK2*.

a, b, Hierarchical clustering of single-cell sequencing data for AML case PAPWIU, in which rows were ordered by clustering (**a**) or by position (**b**). Each row represents one germline SNP and each column is a single cell. Three clusters (11p LOH, Other, and 11p diploid) were detected according to variant allele frequency, ranging from 0.0 (green) to 1.0 (red). The top two rows indicate the cell type (tumour or normal) and *WT1* D447N mutation status. **b**, Variants within the *WT1* locus are highlighted with a blue box. The cluster 'Other' matches the 11p LOH cluster within the *WT1* locus as the samples in this cluster had mono-allelic genotypes at *WT1*, probably caused by a focal deletion. The cluster 'Other' could also be caused by chimeric cells. However, as all cells in this cluster have the same pattern matching, the 11p LOH clusters within the *WT1* gene (the blue box in **b** represents the genomic location of chr11:32,410,002-32,461,785 and *WT1* is located at chr11:32,409,322-32,457,081). A *WT1* focal deletion better explains the profile in 'Other'. **c**, All nine missense *WT1* mutations with DNA and RNA data. The lowest RNA coverage is 16 for *WT1* R445P

in AML case PABLDZ. Five mutations exhibiting allele-specific expression mutations (two-sided Fisher's exact test $P < 0.01$; exact P values also listed for each mutation) are highlighted in blue (grey for $P \geq 0.01$). AML case PABLDZ (*WT1* R445P) had LOH at the *WT1* locus; LOH was present in the predominant clone at the diagnosis and may mask the presence of ASE in a subclone. **d, e**, Two *JAK2* mutations (R683S and D873N) were detected in B-ALL case PAPEWB, in which D873N showed ASE (DNA MAF is 3/38, RNA MAF is 28/74, Fisher's exact test $P < 0.01$). A single-cell sequencing experiment was designed to investigate whether the ASE could be attributed to subclonal CNA undetectable in the bulk tumour. **d**, The 27 germline SNPs in *JAK2* locus were selected along with the two somatic *JAK2* mutations and the other 46 somatic variants. **e**, Heat map of genotype clusters generated from the 64 assays (4 bulk and 60 single cells) passing single-cell sequencing quality control and the original CGI genotype data. The absence of a cluster of mono-allelic genotypes indicates the absence of 9p LOH, which in turn confirms ASE of D873N.



Extended Data Figure 10 | Pathway-centric overview of mutational landscape in paediatric cancers. a, Heat map of somatic mutations in selected pathways across six histotypes. **b,** Pie chart of mutation frequency in selected pathways. The number of samples in the calculation is

indicated for each histotype. An interactive version of the data is available at the ProteinPaint portal (<https://pecan.stjude.org/proteinpaint/study/pan-target>).

Human hippocampal neurogenesis drops sharply in children to undetectable levels in adults

Shawn F. Sorrells^{1,2*}, Mercedes F. Paredes^{1,3*}, Arantxa Cebrian-Silla⁴, Kadellyn Sandoval^{1,3}, Dashi Qi⁵, Kevin W. Kelley¹, David James¹, Simone Mayer^{1,3}, Julia Chang⁶, Kurtis I. Auguste², Edward F. Chang², Antonio J. Gutierrez⁷, Arnold R. Kriegstein^{1,3}, Gary W. Mathern^{8,9}, Michael C. Oldham^{1,2}, Eric J. Huang¹⁰, Jose Manuel Garcia-Verdugo⁴, Zhengang Yang³ & Arturo Alvarez-Buylla^{1,2}

New neurons continue to be generated in the subgranular zone of the dentate gyrus of the adult mammalian hippocampus^{1–5}. This process has been linked to learning and memory, stress and exercise, and is thought to be altered in neurological disease^{6–10}. In humans, some studies have suggested that hundreds of new neurons are added to the adult dentate gyrus every day¹¹, whereas other studies find many fewer putative new neurons^{12–14}. Despite these discrepancies, it is generally believed that the adult human hippocampus continues to generate new neurons. Here we show that a defined population of progenitor cells does not coalesce in the subgranular zone during human fetal or postnatal development. We also find that the number of proliferating progenitors and young neurons in the dentate gyrus declines sharply during the first year of life and only a few isolated young neurons are observed by 7 and 13 years of age. In adult patients with epilepsy and healthy adults (18–77 years; $n = 17$ post-mortem samples from controls; $n = 12$ surgical resection samples from patients with epilepsy), young neurons were not detected in the dentate gyrus. In the monkey (*Macaca mulatta*) hippocampus, proliferation of neurons in the subgranular zone was found in early postnatal life, but this diminished during juvenile development as neurogenesis decreased. We conclude that recruitment of young neurons to the primate hippocampus decreases rapidly during the first years of life, and that neurogenesis in the dentate gyrus does not continue, or is extremely rare, in adult humans. The early decline in hippocampal neurogenesis raises questions about how the function of the dentate gyrus differs between humans and other species in which adult hippocampal neurogenesis is preserved.

We used 59 post-mortem and post-operative samples of the human hippocampus (Supplementary Table 1) to investigate the presence of progenitor cells and young neurons from fetal to adulthood stages. At 14 gestational weeks, at the peak of proliferation in the fetal dentate gyrus (DG)¹⁵, many dividing (Ki-67⁺) neural progenitors (SOX1⁺ (ref. 16) and SOX2⁺ (ref. 17)) were observed in the dentate neuroepithelium (dNE; Fig. 1a, Extended Data Fig. 1a–c and Supplementary Video 1). A continuous region of Ki-67⁺SOX1⁺ and Ki-67⁺SOX2⁺ cells, associated with ribbons of nestin⁺vimentin⁺ fibres and cells, was observed between the dNE and the proximal blade of the DG. At 22 gestational weeks, the proliferating cells between the dNE and the DG were greatly diminished, and most Ki-67⁺SOX1⁺ or Ki-67⁺SOX2⁺ cells in the hippocampus were found in the hilus (Fig. 1b and Extended Data Fig. 1d–f). By this age, most young neurons (DCX⁺PSA-NCAM⁺ cells), were concentrated in the granule cell

layer (GCL) proximal to the dNE (Fig. 1c). By contrast, the distal GCL contained higher numbers of mature NeuN⁺ neurons (Extended Data Fig. 1g, h), suggesting a gradient of maturation.

To look for the formation of a proliferative subgranular zone (SGZ), we characterized dividing and progenitor cells in the human DG from fetal development to adulthood. At 22 gestational weeks, Ki-67⁺ cells were predominantly observed in the hilus and next to the distal GCL (Figs 1b, 2a). By early postnatal life, Ki-67⁺ cells remained distributed throughout the hilus and GCL (Fig. 2a). The number of Ki-67⁺Sox1⁺ or Ki-67⁺Sox2⁺ cells decreased in the hilus during the first year of life (Fig. 2b–d), but these cells did not form a discrete layer beneath the GCL at any of the ages studied (Fig. 2a–d). There were rare instances of SOX2⁺Ki-67⁺ cells in the DG of a 35-year-old individual, but these cells were BLBP[−] and were dispersed throughout the hippocampus. Light and electron microscopy images of samples obtained from individuals at 22 gestational weeks, birth and 7, 18 and 48 years of age did not reveal a layer of cells with progenitor characteristics adjacent to the GCL (Extended Data Figs 2a–c, 3). Ki-67⁺BLBP⁺ cells were found in the developing DG during fetal and early postnatal stages, but BLBP⁺ cells were Ki-67[−] in juvenile and adult brains and were located primarily in the molecular layer (Extended Data Figs 1a, 2b). Furthermore, immunostaining for nestin, vimentin or GFAP in brain sections from individuals that were 7 years of age and older did not show cells next to the DG or in the hilus that had the typical neural progenitor/stem cell morphology of radial astrocytes (also known as radial or type-I cells)^{3,4,18}. BLBP and vimentin were depleted from the hilus and were predominantly expressed in mature stellar astrocytes in the molecular layer in brains from individuals that were 7 years of age and older. GFAP-expressing cells that remained in the adult hilus were stellate ALDH1L1⁺ astrocytes with thin fibres that extended through the hilus and GCL (Extended Data Figs 3a, d, 4). These cells were not Ki-67⁺ or found in mitosis. These results indicate that a germinal SGZ does not form next to the human GCL, and that proliferating cells, which express progenitor/stem cell markers, are mostly depleted from the hilus by 7 years of age.

We next investigated the presence of young neurons in the post-natal, human DG. At birth, DCX⁺PSA-NCAM⁺ cells were located across the GCL, frequently in clusters (Fig. 3a, b). The number of DCX⁺PSA-NCAM⁺ cells in the GCL decreased from 1,618 ± 780 (mean ± s.d.) cells per mm² at birth to 292.9 ± 142.8 cells per mm² at 1 year of age. By 7 years of age, 12.4 ± 5.3 DCX⁺PSA-NCAM⁺ cells per mm² were found in the GCL and at 13 years of age, the GCL contained

¹Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, University of California San Francisco, San Francisco, California 94143, USA. ²Department of Neurological Surgery, University of California San Francisco, San Francisco, California 94143, USA. ³Department of Neurology, University of California San Francisco, San Francisco, California 94143, USA. ⁴Laboratorio de Neurobiología Comparada, Instituto Cavanilles, Universidad de Valencia, CIBERNED, Valencia, 46980, Spain. ⁵State Key Laboratory of Medical Neurobiology and Institutes of Brain Science, Department of Neurology, Zhongshan Hospital, Fudan University, Shanghai, 200032, China. ⁶David Geffen School of Medicine, Department of Neurosurgery, Intellectual Development and Disabilities Research Center, University of California Los Angeles, Los Angeles, California 90095, USA. ⁷Unidad de Cirugía de la Epilepsia, Hospital Universitario La Fe, Valencia 46026, Spain. ⁸Department of Neurosurgery, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California 90095, USA. ⁹Department of Psychiatry and Biobehavioral Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California 90095, USA. ¹⁰Department of Pathology, University of California San Francisco, San Francisco, California 94143, USA.

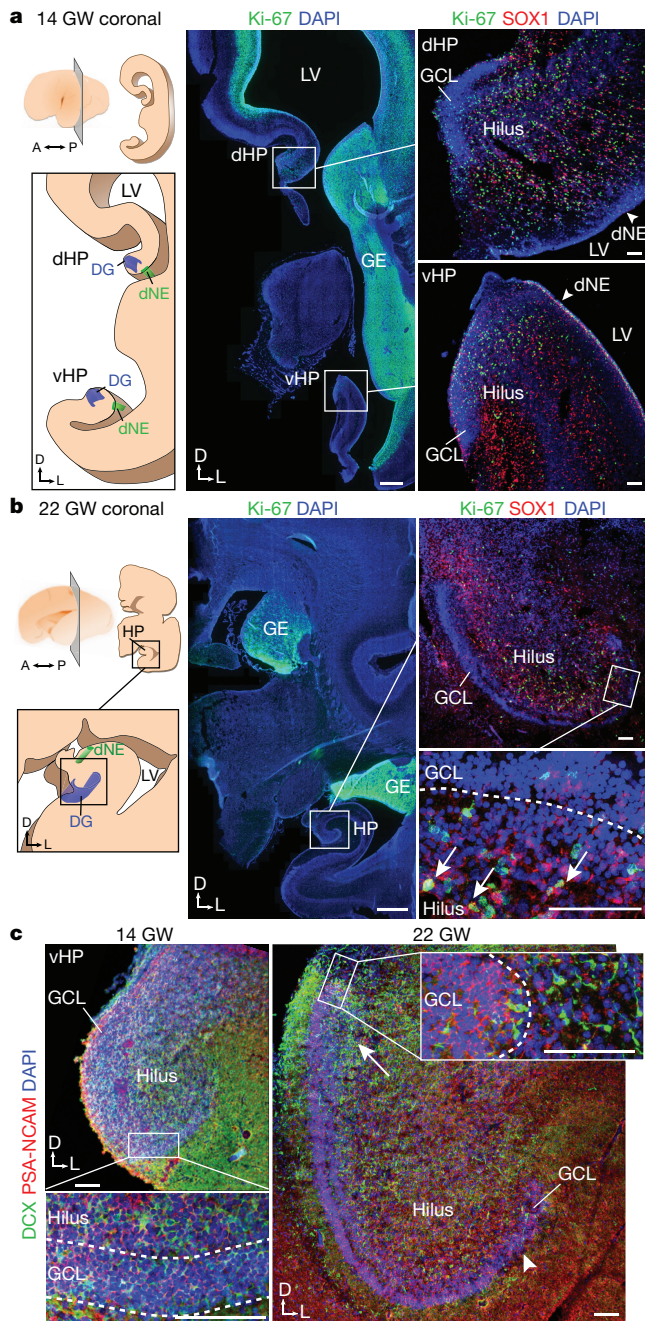


Figure 1 | Early fetal development of the human DG. **a**, Left, schematic of one hemisphere at 14 gestational weeks (GW), coronal section showing the lateral ventricle (LV) and ganglionic eminence (GE). Middle and right, immunofluorescence images. Ki-67⁺SOX1⁺ cells were found in the dorsal (dHP) and ventral (vHP) hippocampus. A, anterior; D, dorsal; L, lateral; P, posterior. **b**, Left, schematic of one hemisphere at 22 gestational weeks, coronal section. Middle and right, SOX1⁺Ki-67⁺ cells in the hilus and GCL (arrows). **c**, Distribution of DCX⁺PSA-NCAM⁺ cells at 14 gestational weeks and 22 gestational weeks. The arrow indicates the end of the GCL most proximal to the dNE. The distal end of the GCL contained fewer young neurons (arrowhead). Scale bars, 1 mm (**a**, **b** middle) and 100 μm (**a**, **b** right), **c**. Staining was replicated at least three times ($n = 1$ at 14 gestational weeks; $n = 3$ at 22 gestational weeks).

2.4 ± 0.74 DCX⁺PSA-NCAM⁺ cells per mm² (that is, approximately 1–2 DCX⁺PSA-NCAM⁺ cells per section; Fig. 3c–e and Extended Data Fig. 5). DCX⁺ cells in the DG of infants (1 year old or less) not only expressed PSA-NCAM, but also frequently had the simple elongated morphology of young neurons (Extended Data Fig. 5b). By contrast,

light and electron microscopy images of sections from the brain of a 7-year-old individual showed that the DG contained DCX⁺ cells in different stages of maturation (Extended Data Fig. 6a). DCX⁺ cells in the hippocampus of a 13-year-old individual had a more mature morphology (Fig. 3e), expressed NeuN and had distinct axons and dendrites (Extended Data Fig. 5c). We examined hippocampi of 17 individuals that were between 18 and 77 years old when they died (Supplementary Table 1) for evidence of young neurons. In two adults (sample numbers 24 and 26), we also studied the ventricular wall and found rare DCX⁺ cells with a migratory morphology in the ventricular–subventricular zone^{19,20}, providing a positive control (Extended Data Fig. 6b). We found no evidence of DCX⁺PSA-NCAM⁺ young neurons in the hilus or GCL of the hippocampus from these individuals (Extended Data Figs 5d, 6b). At three weeks of age, there were many DCX⁺TUJ1⁺ young neurons in the GCL, however, we did not detect these cells at 19 or 36 years of age. In adults, we observed TUJ1⁺ fibres that belonged to many mature neurons (Extended Data Fig. 6c). PSA-NCAM⁺ cells were present in the hilus and GCL of adult brains, but these cells had a mature neuronal morphology and were NeuN⁺ (Extended Data Fig. 5d–f). Using single-molecule *in situ* hybridization labelling of DCX transcripts, we detected many DCX⁺ cells in the GCL at 14 gestational weeks, but only weak signal in very few, widely distributed cells at 13 years (Extended Data Fig. 6d). A subpopulation of cells with round nuclei was occasionally labelled by DCX antibodies. These DCX⁺ cells had multiple processes, were not restricted to the hippocampus, expressed the glial markers IBA1 or OLIG2, and had ultrastructural features of glia (Extended Data Fig. 7).

We also analysed the proliferation of progenitor cells and the presence of young neurons in surgical resections of patients with epilepsy that contained the hippocampus (Supplementary Table 1). In these samples, Ki-67⁺BLBP⁺SOX2⁺ or Ki-67⁺SOX1⁺vimentin⁺ cells were present in the hilus and GCL of a 10-month-old individual, but were absent from the sample of an 11-year-old individual (Extended Data Fig. 8a, b). We also found many DCX⁺PSA-NCAM⁺ cells at 10 months, whereas only a few cells per section were found in samples from a 7-year-old individual and none were found in 13 surgical resections from individuals that were older than 11 years of age (Extended Data Fig. 8c–g). There was no evidence of a discrete layer of dividing cells or young neurons in any of the adult cases with epilepsy that we studied.

We next searched for proliferative progenitors and young neurons in the rhesus macaque (*M. mulatta*). Early studies², in which thymidine-labelling was used, found no evidence of newly generated neurons in adult macaques (17 years old), but subsequent work²¹ using injections of 5-bromodeoxyuridine (BrdU, a thymidine analogue that labels newly generated cells) were used, have suggested that low levels of neurogenesis occur, even in the DG of 23-year-old monkeys. At embryonic day (E)150, we observed remnants of the migratory stream between the dNE and the proximal blade of the developing DG (Extended Data Fig. 9a). Ki-67⁺ and DCX⁺ cells consolidated into a layer in the SGZ between E150 and birth (Fig. 4 and Extended Data Fig. 9a–c). Between birth and 1.5 years of age, the number of Ki-67⁺ cells decreased eightfold and the macaque SGZ became less defined (Fig. 4a). The average number of proliferating cells decreased 35-fold between 1.5 and 7 years of age (Fig. 4e). A continuous SGZ was not detected in macaques that were older than 7 years. Instead, isolated, small, dark cells and occasional Ki-67⁺ cells were observed next to the GCL (Fig. 4a and Extended Data Fig. 9b). Similarly, the number of DCX⁺PSA-NCAM⁺ young neurons decreased during this period, becoming sparse and discontinuous by 7 years of age (Fig. 4b–d, f). Most DCX⁺PSA-NCAM⁺ cells in samples from macaques that were 5 years and older had round nuclei and extensive dendritic trees (Fig. 4c, d and Extended Data Fig. 9d), but some retained the elongated morphology and ultrastructure of young neurons (Fig. 4d, g). Although DCX⁺ cells in the DG of 22- and 23-year-old macaques were rare, they were readily found in the ventricular–subventricular zone and rostral migratory stream²² (Extended Data Fig. 9e). We next used BrdU to

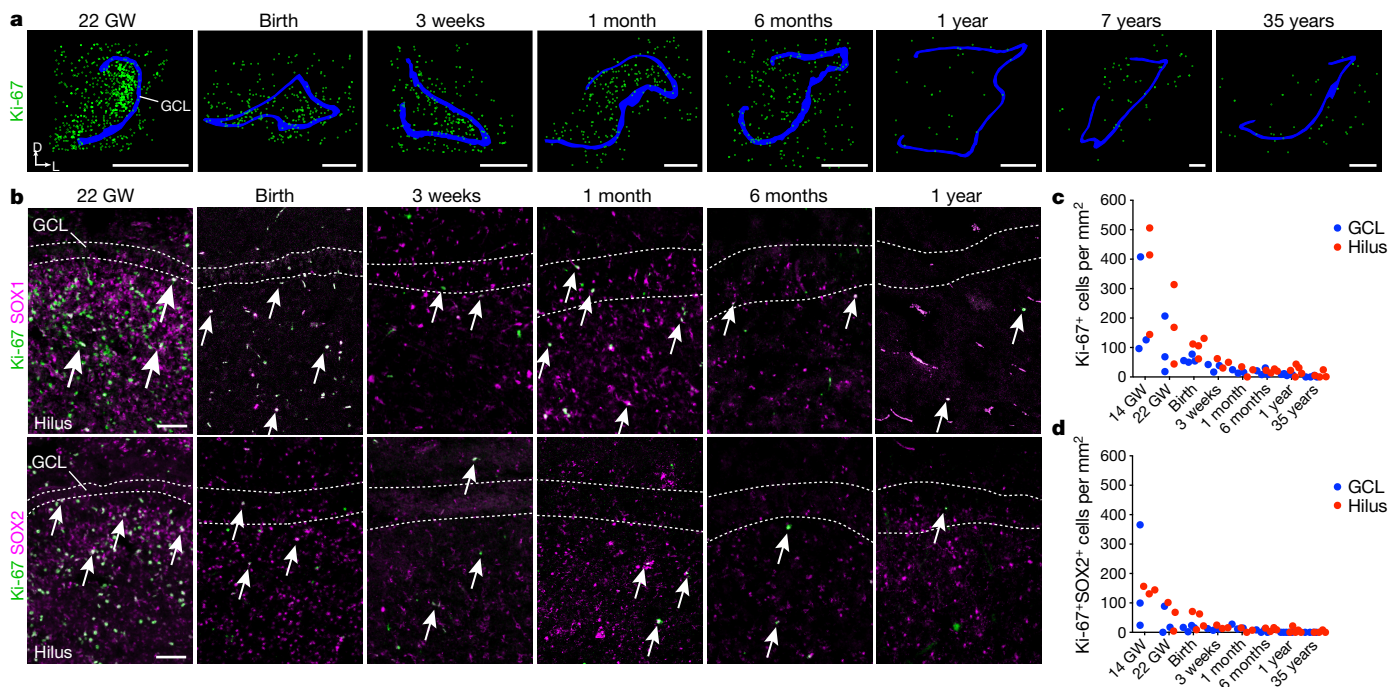


Figure 2 | Human DG proliferation declines sharply during infancy and a layer of proliferating progenitors does not form in the SGZ.
a, Maps of Ki-67⁺ (green) cells in the DG from samples of individuals that were between 22 gestational weeks and 35 years of age; GCL in blue.
b, Ki-67⁺SOX1⁺ and Ki-67⁺SOX2⁺ cells (arrows) are distributed across

the hilus and GCL and the number of double-positive cells decreases between 22 gestational weeks and 1 year of age. **c**, **d**, Quantification of Ki-67⁺ (**c**) and Ki-67⁺SOX2⁺ (**d**) cells in the hilus and GCL. For quantifications, dots indicate staining replicates (≥ 3) (each age $n = 1$). Scale bars, 1 mm (**a**) and 100 μm (**b**).

label recently dividing cells in two 1.5-year-old macaques; at this age the SGZ contained markers of progenitors and young neurons (Extended Data Fig. 9f, g). We checked for BrdU staining 10 and 15 weeks after five days of twice-daily BrdU (50 mg kg⁻¹) injections. DCX⁺BrdU⁺

and a few NeuN⁺BrdU⁺ cells were observed in the SGZ and GCL (Extended Data Fig. 9h, i and Supplementary Table 4). By contrast, in the brains of 7-year-old macaques that received the same BrdU treatment, we found no DCX⁺BrdU⁺ cells in the SGZ 10 weeks after BrdU

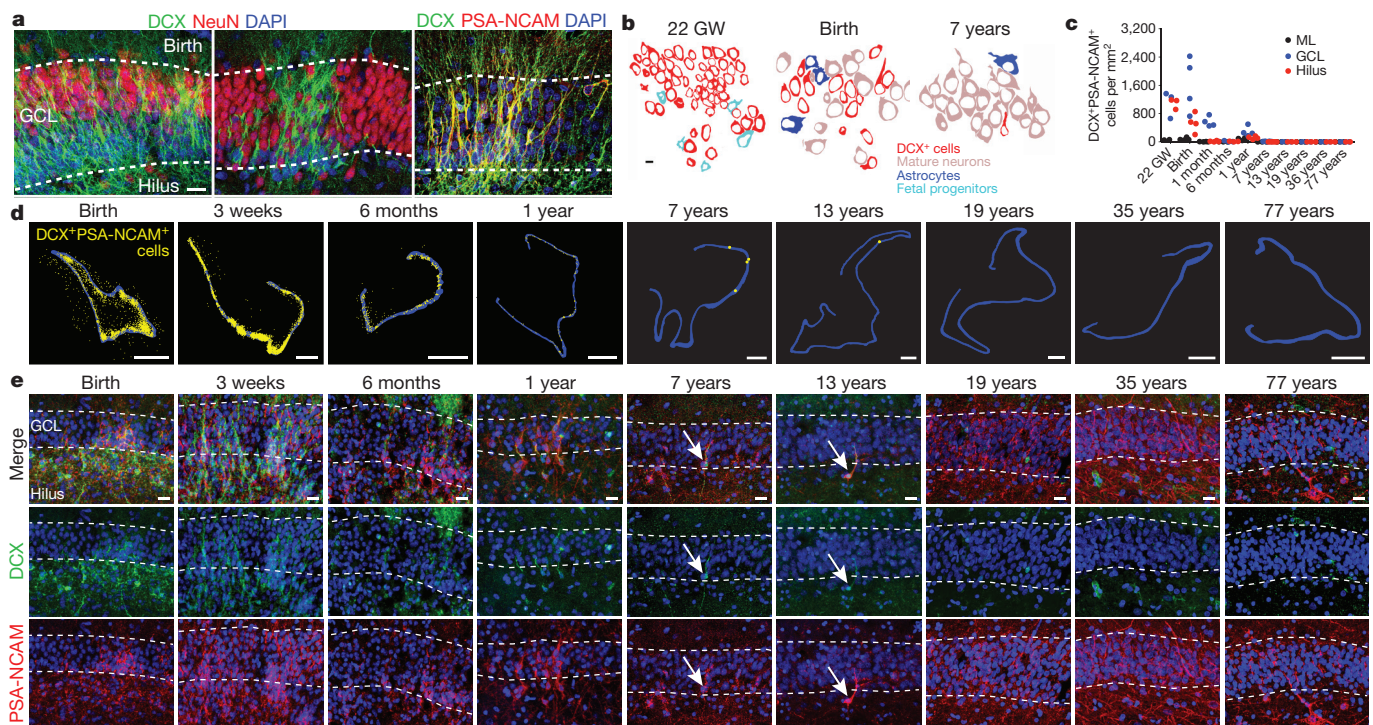


Figure 3 | The number of young neurons declines in the human DG from infancy into childhood. **a**, DCX⁺ cells at birth are distributed in a continuous field (left) or tight clusters (middle) and express PSA-NCAM (right). **b**, Outlines of cell types in the GCL at 22 gestational weeks, birth and 7 years of age. **c**, Quantification of DCX⁺PSA-NCAM⁺ cells in the

DG. **d**, Maps of DCX⁺PSA-NCAM⁺ cells (yellow dots; GCL, blue outline). **e**, DCX⁺PSA-NCAM⁺ cells in the DG (birth to 77 years) are rare by 7 and 13 years of age (arrows). For quantifications, dots indicate staining replicates (≥ 3) (each age, $n = 1$). Scale bars, 1 mm (**d**), 20 μm (**a**, **e**) and 5 μm (**b**).

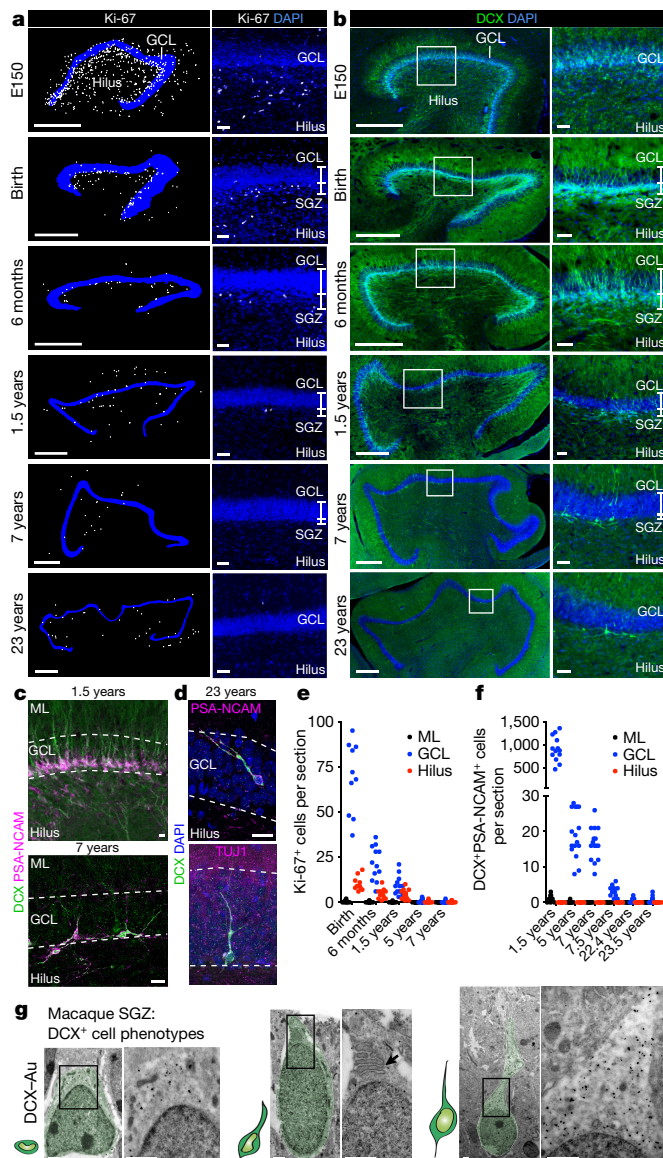


Figure 4 | The SGZ forms during macaque development but new neurons are rare in adults. **a, b,** Maps and immunostaining of Ki-67⁺ cells (**a**) and DCX⁺ cells (**b**) in the macaque SGZ (from E150 to 23 years of age). **c,** DCX⁺PSA-NCAM⁺ cells in the SGZ (1.5 and 7 years). **d,** DCX⁺PSA-NCAM⁺ or DCX⁺TUJ1⁺ cells (23 years). **e, f,** Quantification of Ki-67⁺ cells (**e**) and DCX⁺PSA-NCAM⁺ cells (**f**) in the macaque GCL, hilus and molecular layer (ML). $n = 1$ animal per age; dots indicate staining replicates (≥ 3). **g,** Immunogold (DCX-Au) transmission electron microscopy of neurons (light green overlay) at different stages of maturation. Left, small DCX⁺ cell; middle, DCX⁺ cell with a short process, mitochondria and prominent endoplasmic reticulum (arrow); right, large DCX⁺ cell with round soma, few organelles and an expansion into the GCL. Scale bars, 500 μm (**a, b** (left)), 50 μm (**a, b** (right)), 20 μm (**c, d**) and 1 μm (**g**).

treatment; 15 weeks after BrdU treatment, we found two DCX⁺BrdU⁺ cells (Extended Data Fig. 9j and Supplementary Table 4). We did not find NeuN⁺BrdU⁺ cells in the GCL of these 7-year-old monkeys. Given the higher level of neurogenesis observed in the 1.5-year-old macaque, we studied one monkey at this age 2 h after a single BrdU injection. Many BrdU⁺ cells that expressed the proliferative markers, Ki-67 and MCM2, and the progenitor marker, SOX2, were present in the SGZ (Extended Data Fig. 9h). Finally, we compared hippocampal gene expression profiles from macaque and human (Extended Data Fig. 10). A sharp decrease in DCX, TUJ1 and Ki-67 expression was observed

in both species. In normalized developmental time, the decrease in DCX-expressing cells was accelerated in human compared to macaque (Extended Data Fig. 10). We conclude that there is a marked decrease in neurogenesis in the macaque DG during juvenile development, and that rare DCX⁺PSA-NCAM⁺ young neurons occur in adults.

In the rodent brain, a proliferative SGZ consolidates around postnatal day (P)10^{23,24}, and neural stem cells within this region continue to generate new neurons into adulthood⁴. In the human brain, however, we did not find an equivalent proliferative region at any of the ages analysed. Ki-67⁺ cells were distributed throughout the fetal and infant hilus and GCL. The adult human SGZ was devoid of precursor cells and young neurons, and instead contained many ALDH1L1⁺GFAP⁺ cells. It is intriguing that we found rare examples of SOX2⁺Ki-67⁺ cells in the adult DG, but these cells were not confined to the hilus or GCL and were BLBP[−]. We cannot exclude the possibility that neural stem cells in humans are BLBP[−] or are highly dispersed, but we did not observe DCX⁺PSA-NCAM⁺ cells in these samples. The simplest explanation is that these cells are dividing local glia, many of which are known to express SOX2^{25,26} (Extended Data Fig. 3c). The lack of a coalesced SGZ could explain the absence (or rarity) of DG neurogenesis in the adult human brain.

The above findings do not support the notion that robust adult neurogenesis continues in the human hippocampus (see Supplementary Discussion). ¹⁴C dating on sorted NeuN⁺ nuclei¹¹ has suggested that many new neurons continue to be generated in the adult human hippocampus, with little decline with age, but additional evidence for high levels of progenitors or young neurons was not shown. Interestingly, considerable interindividual variation was observed in this study, and many individual samples had ¹⁴C levels consistent with no, or little, postnatal neuronal addition. Labelled neuronal cells in the GCL in patients that received a low dose of BrdU¹³, could possibly be explained by processes not associated with cell division^{27,28} (Extended Data Fig. 7f). Other groups find a sharp decline with age in proliferation and markers of DG neurogenesis^{12,14,29}, consistent with the above findings. It has been suggested that a few new neurons continue to be produced in adults based on DCX expression detected by PCR or western blot^{14,29,30}. However, glial cells can express DCX²⁶ (Extended Data Fig. 7c–e), possibly explaining some of these expression data. The lack of young neurons in our adult human DG samples could be due to processes linked to disease and/or death. However, similar results were obtained in DG from intraoperative samples or from patients with diverse causes of death. By contrast, young neurons were found in epilepsy samples from children and in our control paediatric cases, despite diverse clinical histories. In contrast to our observations in humans, we observed a germinal SGZ in the young macaques. We found that neurogenesis continues postnatally in macaques, but like humans, this process declined in juveniles and adults, consistent with previous ³H-thymidine and BrdU studies^{2,21,31}. If neurogenesis continues in the adult human hippocampus, this is a rare phenomenon, raising questions of how human DG plasticity differs from other species in which adult hippocampal neurogenesis is abundant. Interestingly, a lack of neurogenesis in the hippocampus has been suggested for aquatic mammals (dolphins, porpoises and whales)⁵, species known for their large brains, longevity and complex behaviour. Understanding the limitations of adult neurogenesis in humans and other species is fundamental to interpreting findings from animal models.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 April 2017; accepted 6 February 2018.

Published online 7 March 2018.

- Altman, J. & Das, G. D. Autoradiographic and histological evidence of postnatal hippocampal neurogenesis in rats. *J. Comp. Neurol.* **124**, 319–335 (1965).
- Kornack, D. R. & Rakic, P. Continuation of neurogenesis in the hippocampus of the adult macaque monkey. *Proc. Natl Acad. Sci. USA* **96**, 5768–5773 (1999).

3. Seri, B., García-Verdugo, J. M., McEwen, B. S. & Alvarez-Buylla, A. Astrocytes give rise to new neurons in the adult mammalian hippocampus. *J. Neurosci.* **21**, 7153–7160 (2001).
4. van Praag, H. *et al.* Functional neurogenesis in the adult hippocampus. *Nature* **415**, 1030–1034 (2002).
5. Patzke, N. *et al.* In contrast to many other mammals, cetaceans have relatively small hippocampi that appear to lack adult neurogenesis. *Brain Struct. Funct.* **220**, 361–383 (2015).
6. Kempermann, G., Kuhn, H. G. & Gage, F. H. More hippocampal neurons in adult mice living in an enriched environment. *Nature* **386**, 493–495 (1997).
7. van Praag, H., Kempermann, G. & Gage, F. H. Running increases cell proliferation and neurogenesis in the adult mouse dentate gyrus. *Nat. Neurosci.* **2**, 266–270 (1999).
8. Lugert, S. *et al.* Quiescent and active hippocampal neural stem cells with distinct morphologies respond selectively to physiological and pathological stimuli and aging. *Cell Stem Cell* **6**, 445–456 (2010).
9. Malberg, J. E. & Duman, R. S. Cell proliferation in adult hippocampus is decreased by inescapable stress: reversal by fluoxetine treatment. *Neuropsychopharmacology* **28**, 1562–1571 (2003).
10. Hill, A. S., Sahay, A. & Hen, R. Increasing adult hippocampal neurogenesis is sufficient to reduce anxiety and depression-like behaviors. *Neuropsychopharmacology* **40**, 2368–2378 (2015).
11. Spalding, K. L. *et al.* Dynamics of hippocampal neurogenesis in adult humans. *Cell* **153**, 1219–1227 (2013).
12. Dennis, C. V., Suh, L. S., Rodriguez, M. L., Kril, J. J. & Sutherland, G. T. Human adult neurogenesis across the ages: an immunohistochemical study. *Neuropathol. Appl. Neurobiol.* **42**, 621–638 (2016).
13. Eriksson, P. S. *et al.* Neurogenesis in the adult human hippocampus. *Nat. Med.* **4**, 1313–1317 (1998).
14. Knoth, R. *et al.* Murine features of neurogenesis in the human hippocampus across the lifespan from 0 to 100 years. *PLoS ONE* **5**, e8809 (2010).
15. Yang, P. *et al.* Developmental profile of neurogenesis in prenatal human hippocampus: an immunohistochemical study. *Int. J. Dev. Neurosci.* **38**, 1–9 (2014).
16. Venero, M. *et al.* Sox1 marks an activated neural stem/progenitor cell in the hippocampus. *Development* **139**, 3938–3949 (2012).
17. Suh, H. *et al.* *In vivo* fate analysis reveals the multipotent and self-renewal capacities of Sox2⁺ neural stem cells in the adult hippocampus. *Cell Stem Cell* **1**, 515–528 (2007).
18. Steiner, B. *et al.* Type-2 cells as link between glial and neuronal lineage in adult hippocampal neurogenesis. *Glia* **54**, 805–814 (2006).
19. Sanai, N. *et al.* Corridors of migrating neurons in the human brain and their decline during infancy. *Nature* **478**, 382–386 (2011).
20. Wang, C. *et al.* Identification and characterization of neuroblasts in the subventricular zone and rostral migratory stream of the adult human brain. *Cell Res.* **21**, 1534–1550 (2011).
21. Gould, E. *et al.* Hippocampal neurogenesis in adult Old World primates. *Proc. Natl Acad. Sci. USA* **96**, 5263–5267 (1999).
22. Kornack, D. R. & Rakic, P. The generation, migration, and differentiation of olfactory neurons in the adult primate brain. *Proc. Natl Acad. Sci. USA* **98**, 4752–4757 (2001).
23. Sugiyama, T., Osumi, N. & Katsuyama, Y. The germinal matrices in the developing dentate gyrus are composed of neuronal progenitors at distinct differentiation stages. *Dev. Dyn.* **242**, 1442–1453 (2013).
24. Altman, J. & Bayer, S. A. Mosaic organization of the hippocampal neuroepithelium and the multiple germinal sources of dentate granule cells. *J. Comp. Neurol.* **301**, 325–342 (1990).
25. Komitova, M. & Eriksson, P. S. Sox-2 is expressed by neural progenitors and astroglia in the adult rat brain. *Neurosci. Lett.* **369**, 24–27 (2004).
26. Zhang, Y. *et al.* An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J. Neurosci.* **34**, 11929–11947 (2014).
27. Breunig, J. J., Arellano, J. I., Macklis, J. D. & Rakic, P. Everything that glitters isn't gold: a critical review of postnatal neural precursor analyses. *Cell Stem Cell* **1**, 612–627 (2007).
28. Gould, E. How widespread is adult neurogenesis in mammals? *Nat. Rev. Neurosci.* **8**, 481–488 (2007).
29. Mathews, K. J. *et al.* Evidence for reduced neurogenesis in the aging human hippocampus despite stable stem cell markers. *Aging Cell* **16**, 1195–1199 (2017).
30. Fahrner, A. *et al.* Granule cell dispersion is not accompanied by enhanced neurogenesis in temporal lobe epilepsy patients. *Exp. Neurol.* **203**, 320–332 (2007).
31. Eckenhoff, M. F. & Rakic, P. Nature and fate of proliferative cells in the hippocampal dentate gyrus during the life span of the rhesus monkey. *J. Neurosci.* **8**, 2729–2747 (1988).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the families who donated the tissue samples used in this study, and J. Rodriguez, V. Tang, J. Cotter and C. Guinto for technical support. S.F.S. was supported by F32 MH103003 and M.F.P. was supported by K08 NS091537. A.A.-B. was supported by NIH grants P01 NS083513, R01 NS028478 and a gift from the John G. Bowes Research Fund. He is the Heather and Melanie Muss Endowed Chair and Professor of Neurological Surgery at UCSF and is a co-founder and serves on the scientific advisory board of Neurona Therapeutics. G.W.M. was partly supported by the Davies/Crandall Endowed Chair For Epilepsy Research at UCLA. G.W.M. and J.C. were supported by NIH NINDS (NS083823 and U01 MH108898). M.C.O. was supported by a Scholar Award from the UCSF Weill Institute for Neurosciences. We acknowledge NSFC grants to Z.Y. (31425011, 31630032, and 31421091). S.M. was supported by fellowships from the European Molecular Biology Organization (EMBO Long-Term Fellowship, ALTF_393-2015) and the German Research Foundation (DFG, MA 7374/1-1). J.M.G.-V. and A.C.-S. were supported by MINECO/FEDER Grant BFU2015-64207-P, Red de Terapia Celular TerCel, Instituto de Salud Carlos III (ISCIII2012-RED-19-016 and RD12/0019/0028) and PROMETEOII/2014/075.

Author Contributions M.F.P. and S.F.S. contributed equally as co-first authors and A.C.-S., K.S. and D.Q. contributed equally as second authors. Z.Y., A.A.-B., M.F.P. and S.F.S. conceived the study. A.A.-B., S.F.S. and M.F.P. designed and interpreted the experiments and, with A.C.-S., K.S., D.Q., S.M. and D.J., conducted the experiments. K.I.A., E.F.C., J.C., E.J.H., A.J.G., A.R.K. and G.W.M. assisted with specimen collection and conducted clinical and neuropathological reviews. K.W.K. and M.C.O. designed and performed the bioinformatic analyses. A.C.-S. and J.M.G.-V. conducted and interpreted the ultrastructural studies. S.F.S., M.F.P., A.C.-S. and K.S. prepared the figures. A.A.-B., S.F.S. and M.F.P. wrote the manuscript with input from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to A.A.-B. (AlvarezBuyllaA@ucsf.edu), Z.Y. (yangz@fudan.edu.cn) or J.G.-V. (j.manuel.garcia@uv.es).

METHODS

Human tissue collection. Thirty-seven post-mortem specimens from controls and twenty-two post-operative neurosurgical specimens from patients with epilepsy were collected for this study (Supplementary Table 1). Tissue was collected at the following institutions with previous patient consent in strict observance of the legal and institutional ethical regulations of each participating institution: (1) The University of California, San Francisco (UCSF) Committee on Human Research. Protocols were approved by the Human Gamete, Embryo and Stem Cell Research Committee (Institutional Review Board) at UCSF. (2) The Ethical Committee for Biomedical Investigation, Hospital La Fe (2015/0447) and the University of Valencia Ethical Commission for Human Investigation. (3) In accordance with institutional guidelines and study design approval by the Institutional Review Board (Ethics Committee) of Shanghai Medical College (20110307-085, 20120302-099). (4) Specimens collected at UCLA had Institutional Review Board-approved informed consents for research along with HIPAA authorizations signed by parents or responsible guardians, as per the UCLA Human Research Protection Program. For infant cases, when the brain was at full term (37–40 gestational weeks) and autopsy was performed within two days after birth, we refer to the case as 'birth'. We collected tissue blocks from the temporal lobe, posteriorly from the amygdaloid complex to the posterior end of the inferior horn of the lateral ventricle. Autopsy samples had a post-mortem interval of less than 48 h, and samples from patients with epilepsy had less than 1 h to fixation (in 4% paraformaldehyde (PFA) or 10% formalin). For two adult brains (35-years old and 39-years old), the individuals were perfused within 3–5 h of death with 4% PFA during autopsy via the carotid artery and placed in fixative. All brains were typically cut into approximately 1.5-cm blocks, fixed in 4% PFA for an additional two days, cryoprotected in a 30% sucrose solution, and then frozen in embedding medium (OCT). Blocks were cut into 30- μ m sections on a cryostat and mounted on glass slides for immunohistochemistry. For each case, we cresyl-stained a minimum of three sections at different levels to confirm anatomical landmarks and orientation of the sample. Neurosurgical excisions of the temporal lobe, which included the hippocampus, were performed as part of resection treatment in patients with intractable epilepsy as previously described³². We recorded the anatomical origin of each intra-operative specimen with intra-operative neuronavigation. Intra-operative specimen assessments were independently confirmed by a neuropathologist. We performed immunohistochemistry staining on surgical sections to look for the expression of PROX1 to confirm the location of the GCL.

Macaque tissue preparation. All animal care and experiments were conducted in accordance with the Fudan University Shanghai Medical College and UC Davis guidelines. Embryonic, neonatal, juvenile and adult macaque monkeys, *M. mulatta*, of both sexes at various ages (Supplementary Table 2), were obtained from the Kunming Primate Research Center of the Chinese Academy of Sciences (Kunming, China), Suzhou Xishan Zhongke Laboratory Animal Co., Ltd (Suzhou, China) and the UC Davis Primate Research Center (Davis, USA). For immunohistochemical staining, postnatal monkeys were deeply anaesthetized and then perfused with PBS followed by 4% PFA. The brains were removed and post-fixed with 4% PFA for 12–48 h. Postnatal brains were then cut coronally into approximately 1.0–2.0-cm slabs and cryoprotected in 30% sucrose in 0.1 M phosphate buffer at 4°C for 72 h. The brain tissue samples were frozen in embedding medium (OCT) on a dry ice and ethanol slush.

BrdU administration. We used five monkeys to do BrdU labelling experiments: three 1.5-year-old monkeys and two 7–7.5-year-old monkeys. BrdU acute labelling: one 1.5-year-old monkey was injected once intravenously with BrdU (50 mg kg⁻¹) and euthanized 2 h after BrdU injection. BrdU birth dating: BrdU (50 mg kg⁻¹) was injected intravenously twice daily for five days in two 1.5-year-old and two 7–7.5-year-old monkeys. One 1.5-year-old monkey and the 7.5-year-old monkey were euthanized 10 weeks after BrdU injections; another 1.5-year-old monkey and the 7-year-old monkey were euthanized 15 weeks after BrdU injections. We analysed 52 sections for the presence of BrdU labelling in the brain of the 7.5-year-old macaque after a 10-week delay and 76 sections of the brain of the 7-year-old after a 15-week delay.

Immunohistochemistry. Frozen slides were allowed to equilibrate to room temperature for 3 h. Some antigens required antigen retrieval (Supplementary Table 3), which was conducted at 95°C in 10 mM sodium citrate buffer, pH 6.0. Following antigen retrieval, slides were washed with TNT buffer (0.05% Triton-X100 in PBS) for 10 min, placed in 1% H₂O₂ in PBS for 45 min and then blocked with TNB solution (0.1 M Tris-HCl, pH 7.5, 0.15 M NaCl, 0.5% blocking reagent from PerkinElmer) for 1 h. Slides were incubated in primary antibodies overnight at 4°C (Supplementary Table 3) and in biotinylated secondary antibodies (Jackson ImmunoResearch Laboratories) for 2.5 h at room temperature. All antibodies were diluted in TNB solution. For most antibodies, the conditions of use were validated

by the manufacturer (antibody product sheets). When this information was not provided, we performed control experiments, including no primary antibody (negative) controls and comparison to mouse staining patterns.

Sections were then incubated for 30 min in streptavidin–horseradish peroxidase, which was diluted (1:200) with TNB. Tyramide signal amplification (PerkinElmer) was used for some antigens. Sections were incubated in tyramide-conjugated fluorophores for 5 min at the following dilutions: fluorescein: 1:50; Cy3: 1:100; Cy5: 1:100. For sections that used the 3',3'-diaminobenzidine (DAB) chromogenic immunohistochemistry method, the sections were first rinsed in PBS, incubated for 15 min in 1% H₂O₂, then incubated for 2 h in 10% fetal calf serum as the blocking buffer. This was followed by overnight incubation with the primary antibody at 4°C, followed by incubation with the secondary antibody for 2 h at room temperature, and development using the VECTASTAIN ABC HRP system (Vector Laboratories). After several PBS rinses, sections were dehydrated, mounted and coverslipped. Staining was conducted in technical triplicates before analysis.

Fluorescent microscopy, image processing and quantifications. Images were acquired on Leica TCS SP8 or SP5 confocal microscopes using 10 \times (0.3 NA) or 63 \times (1.4 NA) objective lenses. Imaging of entire sections and for quantification of DCX⁺PSA-NCAM⁺ cells was carried out at 20 \times (0.45 NA) magnification on a Zeiss Axiovert 200M microscope or Keyence BZ-X Analyzer (BZX700) and individual files were stitched automatically. Imaging files were analysed and quantified in Neurolucida software (MBF Bioscience, 2017 version). Linear adjustments to image brightness and contrast were made equivalently across all images using Adobe Photoshop (CS 6). Cells were counted in Z-stack images from sections stained with Ki-67 and SOX2 or DCX and PSA-NCAM. Three to five representative images across a minimum of three evenly spaced and randomly sampled sections were collected for quantification at each age. Experimental replicates and different co-stains (in addition to the 3–5 sections included for quantifications) were also analysed for the presence or absence of young neurons or stem cells. The DG was subdivided into regions of interest (GCL, hilus or molecular layer) using DAPI to initially identify the cell-dense GCL. Each age has $n = 1$. Counts for cell populations were performed by three separate investigators who were blinded to individual cases. For each quantified marker, counts were repeated by different investigators for reproducibility. Fluorescence signal for single reactivity and co-localization of immunoreactivity was counted individually using the markers function in the Neurolucida imaging software. The quantification of data was performed with GraphPad Prism (v.6). No statistical methods were used to predetermine sample size.

Electron microscopy. For transmission electron microscopy (TEM), samples were sectioned with a vibrating blade microtome (200 μ m) and post-fixed with 2% osmium tetroxide solution. Sections were dehydrated in increasing ethanol concentrations and stained with 2% uranyl acetate, embedded in araldite resin (Durcupan ACM Fluka, Sigma-Aldrich), and allowed to solidify at 69°C for 72 h. We analysed 6 controls and 15 cases with epilepsy via TEM. We looked for the presence of cell clusters under light and electron microscopy in the 15 resected samples from patients with epilepsy that were 30–64 years old (at least 15 semi-thin sections per case) and 4 control samples from individuals that were 18–55 years old (45 semi-thin sections per case). In the additional control cases (samples from individuals that were 7 years old and 48 years old), we studied 100 semi-thin sections spanning the entirety of the anterior to posterior levels of the DG. Ultrathin sections were obtained (70 nm) for all controls and cases, and were contrasted with lead citrate solution on grids. Pre-embedding immunohistochemistry was performed on 50- μ m floating sections with DCX and IBA1 antibodies. Post-fixation was performed with 7% glucose–1% osmium tetroxide, after which a conventional embedding protocol was followed. TEM micrographs of DCX immunolabelled ultrathin sections of the DG from individuals that were obtained at 22 gestational weeks of age (proximal edge), birth and 7 years of age were used for the GCL cellular profiles. All images were taken at the same magnification. Cell profiles were drawn on Adobe Photoshop by following the cytoplasmic cell membranes. Cells showing DCX immunogold labelling were coloured in red. DCX⁺ cells were identified by their ultrastructural characteristics: progenitors (light blue) had dark cytoplasm and few intermediate filaments and ensheathed DCX⁺ cells; astrocytes (blue) had an irregular contour, star-shape morphology, light cytoplasm and intermediate filaments; mature neurons showed a large cell body with a large, round nucleus, and high amounts of ribosomes and organelles.

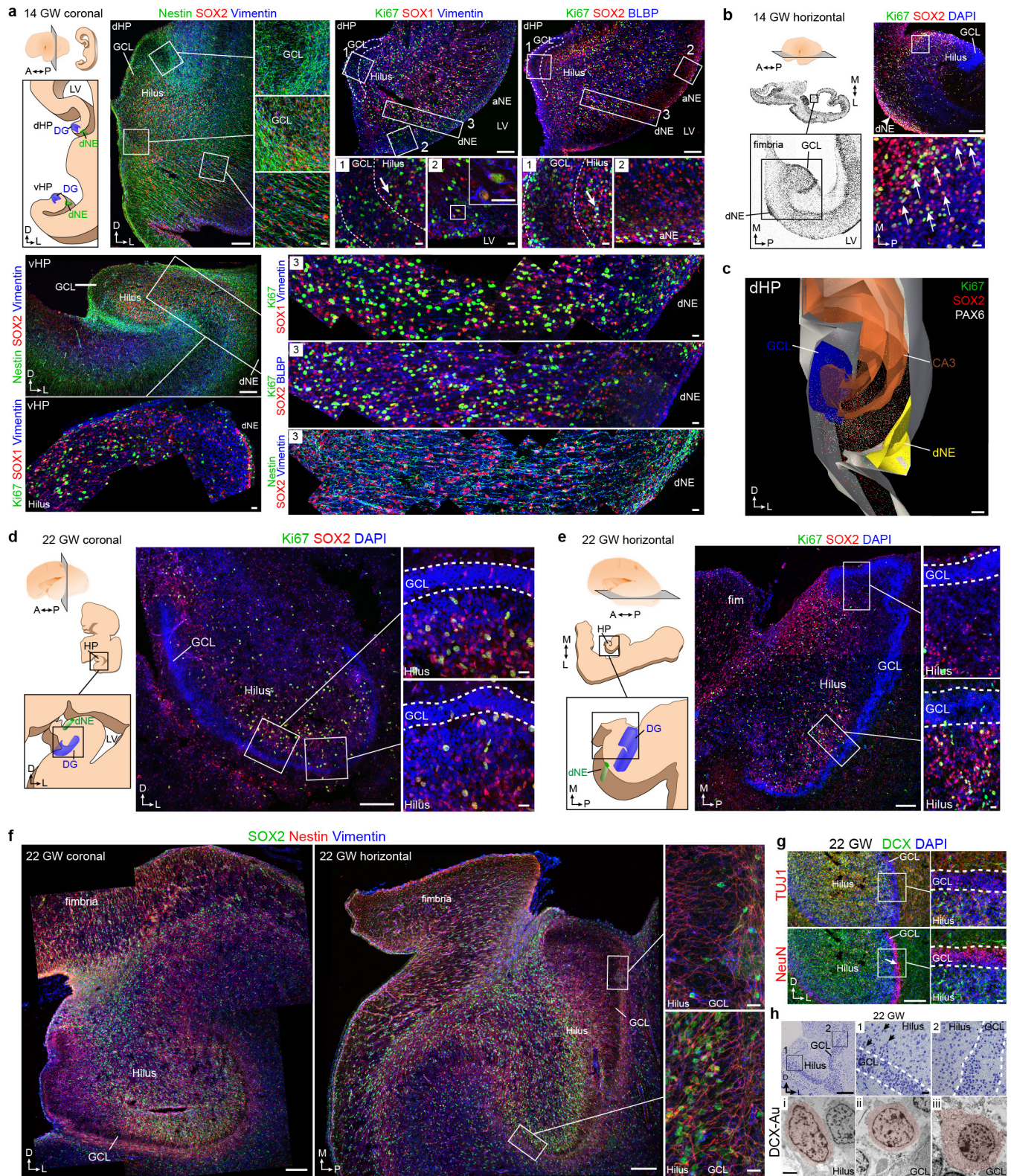
RNA scope in situ hybridization. Sequences of target probes, preamplifier, amplifier and label probes are proprietary and commercially available (Advanced Cell Diagnostics). Typically, the probes contain 20 ZZ probe pairs (approximately 50 bp per pair) covering around 1,000 bp. Here, we used a probe against human DCX targeting 181–1381 of NM_000555.3 as a single-plex probe. Slides for *in situ* hybridization were initially taken from –80°C and dried at 60°C for 1 h and fixed in 4% PFA for 2 h. After several PBS washes, slides were treated with ACD hydrogen peroxide for 10 min and then washed in water 2 \times before treatment in

1× target retrieval buffer (ACD) for 5 min (at 95–100 °C). After washing in water and then 100% alcohol, the slides were left to dry overnight before protease treatment for 15 min at 40 °C in the RNAscope oven. Hybridization of probes and amplification solutions was performed according to the manufacturer's instructions. In brief, tissue sections were incubated in the desired probe (around 2–3 drops per section) for 2 h at 40 °C. The slides were washed twice in 1× wash buffer (ACD) for 2 min each. Amplification and detection steps were performed using the RNAscope 2.5 HD Red Detection Kit reagents (ACD, 320497) for single-plex probes. Sections were incubated with buffer Amp1 for 30 min at 40 °C and then washed twice in wash buffer for 2 min each. Amp2 was incubated on the sections for 15 min at 40 °C, followed by two washes in wash buffer. Sections were incubated in Amp3 for 30 min at 40 °C and washed twice in wash buffer for 2 min each, followed by incubation with Amp4 for 15 min at 40 °C. Slides were washed twice in wash buffer for 2 min each. Slides were incubated with Amp5 for 30 min at room temperature using the HybEZ humidity control tray and slide rack to maintain humidity. The slides were washed twice in 1× wash buffer for 2 min each and incubated in Amp6 for 15 min at room temperature before washing twice in wash buffer for 2 min each. The *in situ* hybridization signal was detected by diluting Fast RED-B in Fast RED-A solution (1:60 ratio) and incubating sections in this solution for 10 min. Slides were washed twice in water to stop the reaction.

Comparative gene transcription analysis. Developmental expression data were downloaded for human hippocampus (<http://brainspan.org/>; RPKM data; October 2013 release) and rhesus macaque hippocampus (<http://blueprintnpatlas.org/>; March 2014 release). To compare laser-capture microdissected rhesus macaque samples to gross human hippocampus samples, we calculated average expression over all hippocampus samples for each age^{33,34}. Expression data were z-score normalized for each species and ages were aligned between species based on calculated event scores of conserved timing of neurodevelopmental events³⁵.

Data availability. All data and/or analyses generated during the current study are available from the corresponding author upon reasonable request.

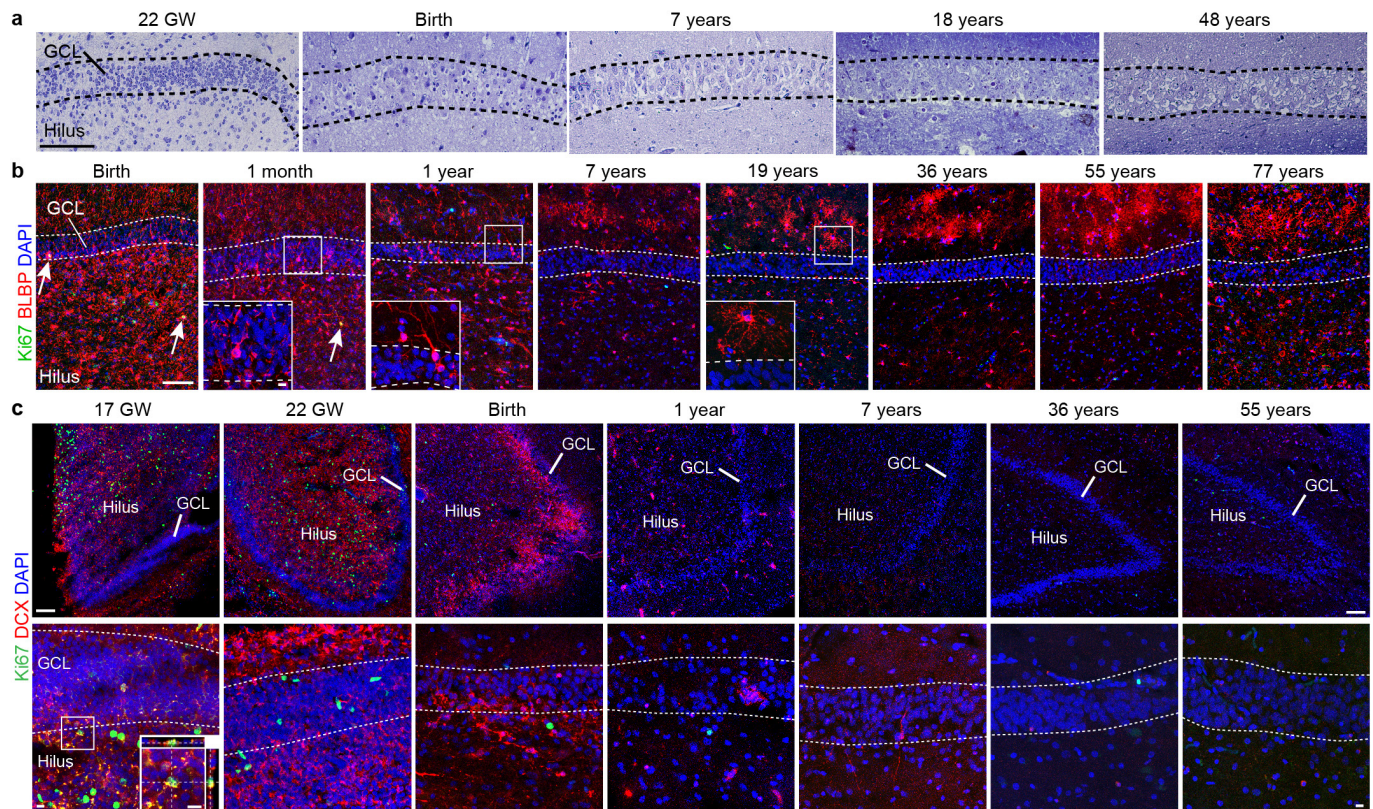
32. Mathern, G. W. *et al.* Seizures decrease postnatal neurogenesis and granule cell development in the human fascia dentata. *Epilepsia* **43**, 68–73 (2002).
33. Bakken, T. E. *et al.* A comprehensive transcriptional map of primate brain development. *Nature* **535**, 367–375 (2016).
34. Kang, H. J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011).
35. Workman, A. D., Charvet, C. J., Clancy, B., Darlington, R. B. & Finlay, B. L. Modeling transformations of neurodevelopmental sequences across mammalian species. *J. Neurosci.* **33**, 7368–7383 (2013).



Extended Data Figure 1 | See next page for caption.

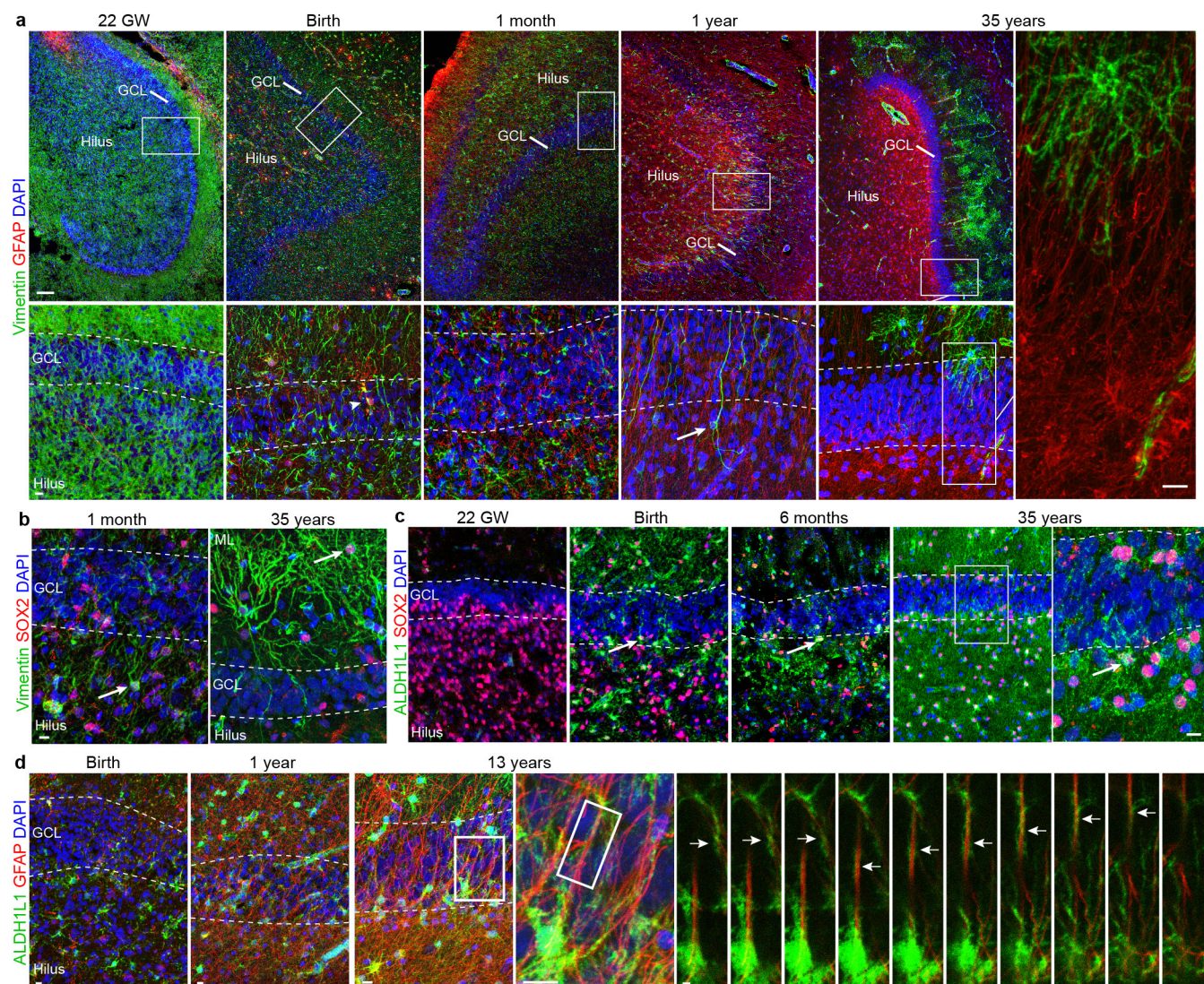
Extended Data Figure 1 | Additional marker and ultrastructural analysis of early fetal development of the human DG. **a–c**, Human brain of an individual at 14 gestational weeks. **a**, Schematic of the dorsal (dHP) and ventral (vHP) hippocampus in a coronal section. Precursor cells labelled with nestin, SOX2 and vimentin are organized in ribbons between the dNE and GCL. Ki-67⁺ cells expressing SOX1 and vimentin or SOX2 and BLBP are present in the GCL and hilus (inset 1), along the wall of the lateral ventricle (LV) (inset 2) and between the GCL and the dNE (inset 3). The dNE is located at the edge of the ammonic neuroepithelium (aNE) closest to the fimbria. A similar organization is present in the vHP where nestin⁺SOX2⁺vimentin⁺ cells connect the dNE to the developing GCL. Ki-67⁺SOX1⁺vimentin⁺ cells are present in a strip along the ventricular wall and fill the region between the dNE and the GCL. **b**, Left, hemisphere at 14 gestational weeks, Nissl-stained horizontal sections. Right, Ki-67⁺ cells expressing SOX2 (arrows). **c**, 3D reconstruction of the dHP showing the field of Ki-67⁺ and SOX2⁺ cells between the dNE and GCL. **d–h**, Human brain at 22 gestational weeks, coronal (**d**) and horizontal (**e**) sections. The hilus and GCL contain Ki-67⁺SOX2⁺ cells (**d**, **e** (insets)) as well as nestin⁺SOX2⁺vimentin⁺ cells (**f**). These

populations are asymmetrically distributed; sparse in the medial (proximal) GCL and hilus (top insets in **e**, **f**) but abundant in the lateral (distal) GCL and hilus (bottom insets in **e**, **f**). **g**, DCX⁺TUJ1⁺ cells and NeuN⁺ cells in the DG at 22 gestational weeks. NeuN⁺ GCL neurons in the distal GCL (arrow). **h**, A toluidine-blue-stained semi-thin section (top) and TEM micrographs showing the ultrastructural characteristics of DCX immunogold-labelled cells (pseudocoloured, bottom) at 22 gestational weeks. Insets of the semi-thin section show the proximal (1) and distal (2) ends of the GCL. Most DCX⁺ cells in the hilus and the proximal GCL have little cytoplasm, few organelles and a small, irregular nucleus (i, ii); some DCX⁺ cells in the hilus have an elongated, fusiform morphology (i). Some DCX⁺ cells in the GCL have mature neuronal characteristics such as a round nucleus, more cytoplasm, ribosomes, rough endoplasmic reticulum and mitochondria (iii); this cell type was more common in the distal GCL. At this stage, the round and more mature neuronal morphologies were observed in the distal, but not in the proximal, blade. Scale bars, 200 μ m (**a–h** (left images)), 2 μ m (**a–h** (insets)) and 2 μ m (**h** (TEM)).



Extended Data Figure 2 | A coalesced proliferative SGZ does not form in the human DG; additional marker expression. **a**, Toluidine-blue-counterstained semi-thin sections of the human GCL from fetal to adult ages. Note that a discrete cellular layer does not form next to the GCL and the small dark cells characteristic of SGZ precursors are not present (compare to Extended Data Fig. 9b in the macaque). **b**, BLBP⁺ cells are distributed broadly in the DG from birth to 1 year, many of these cells have a radial morphology (see insets) and some co-express Ki-67 at birth and

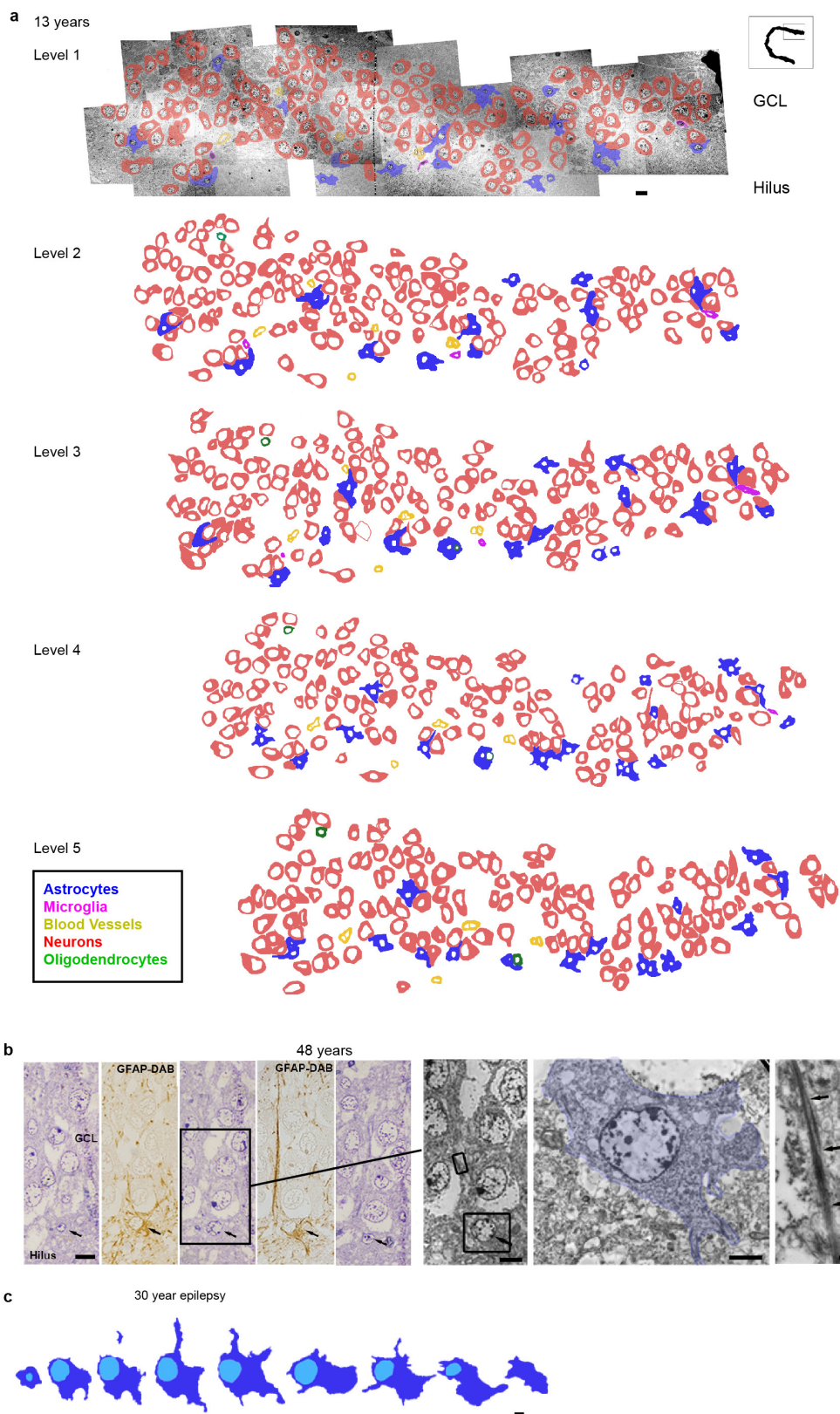
1 month (double-positive cells are indicated by the arrows). By 7 years and in adults, most BLBP is present in the molecular layer in stellate protoplasmic astrocytes. **c**, DCX⁺Ki-67⁺ cells in the GCL are rare at 17 gestational weeks (orthogonal views, inset) but were abundant in the ganglionic eminence at the same age (data not shown). DCX⁺Ki-67⁺ cells were absent in the GCL from 22 gestational weeks to 55 years. Scale bars, 100 μ m (**a–c**) and 10 μ m (**a–c** (insets)).



Extended Data Figure 3 | Additional marker expression for astroglial cells and progenitor cells in the human DG at different ages.

a, Vimentin⁺ and GFAP⁺ cells in the hippocampus from 22 gestational weeks to 35 years. Vimentin is widely expressed during fetal and early postnatal development and is mostly restricted to protoplasmic astrocytes in the molecular layer in adults. GFAP is not expressed at 22 gestational weeks, but at birth a few vimentin⁺GFAP⁺ cells are present in the hilus and GCL (arrowhead). Interestingly, some vimentin⁺GFAP⁺ cells with a radial morphology (arrow) are observed in samples at 1 year of age, but not at the other ages. In adults, GFAP and vimentin are not co-expressed

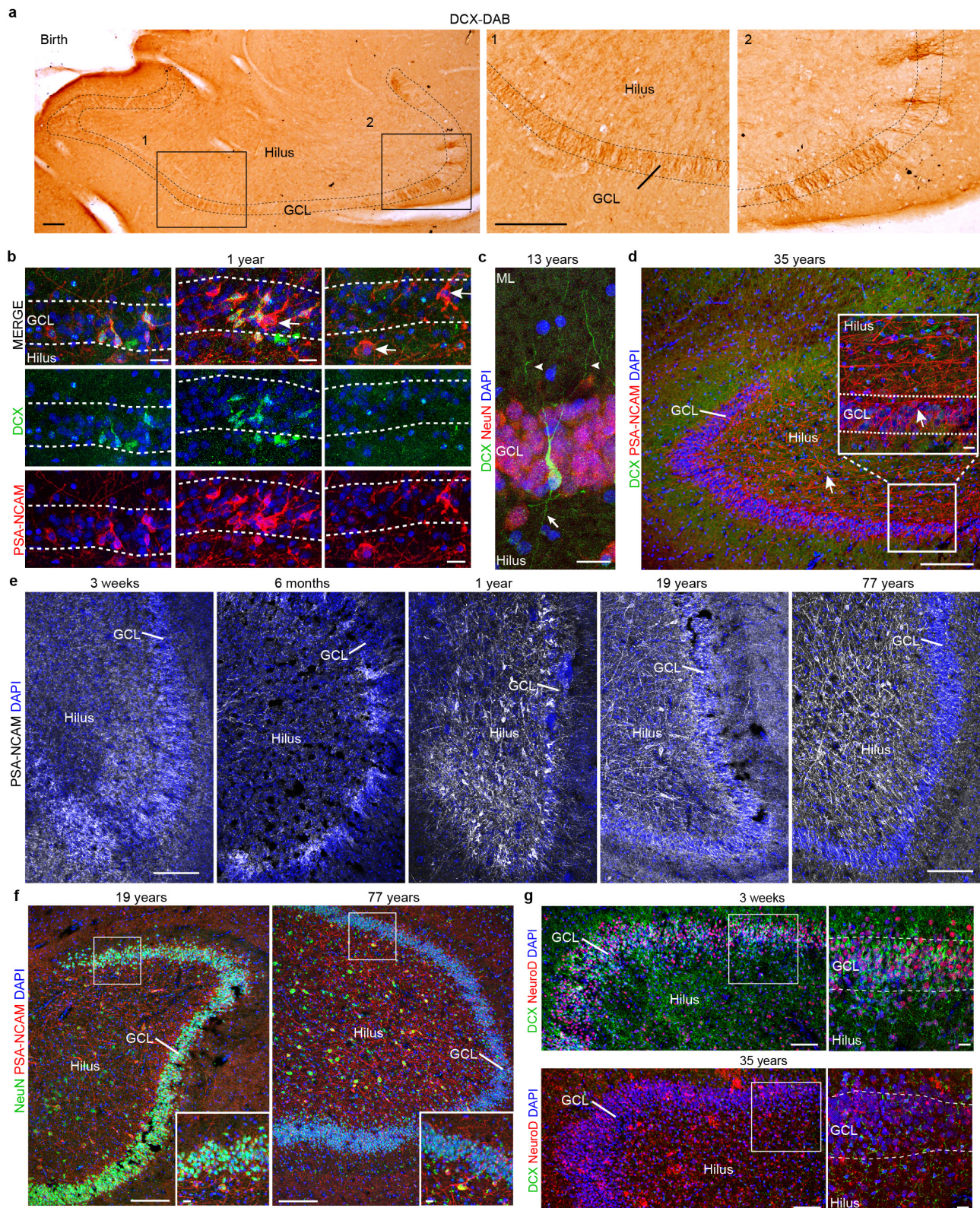
(right, high-magnification of thin GFAP⁺vimentin⁺ fibres within the GCL). **b**, Vimentin⁺Sox2⁺ simpler elongated cells in the hilus at 1 month (arrow) and protoplasmic astrocytes in the molecular layer at 35 years of age (arrow). **c**, SOX2⁺ cells are abundant in the GCL and hilus at 22 gestational weeks, and co-express ALDH1L1 in the brain at birth and in older individuals (arrows). **d**, At birth, there are few ALDH1L1⁺GFAP⁺ cells in the DG, but by 13 years of age many stellate astrocytes express both of these markers. Right, z stack of radial GFAP⁺ processes that are surrounded by ALDH1L1 staining. Scale bars, 100 μ m (**a** (top row)), 10 μ m (**a** (bottom row and insets), **b–d**) and 2 μ m (**d** (z-stack)).



Extended Data Figure 4 | See next page for caption.

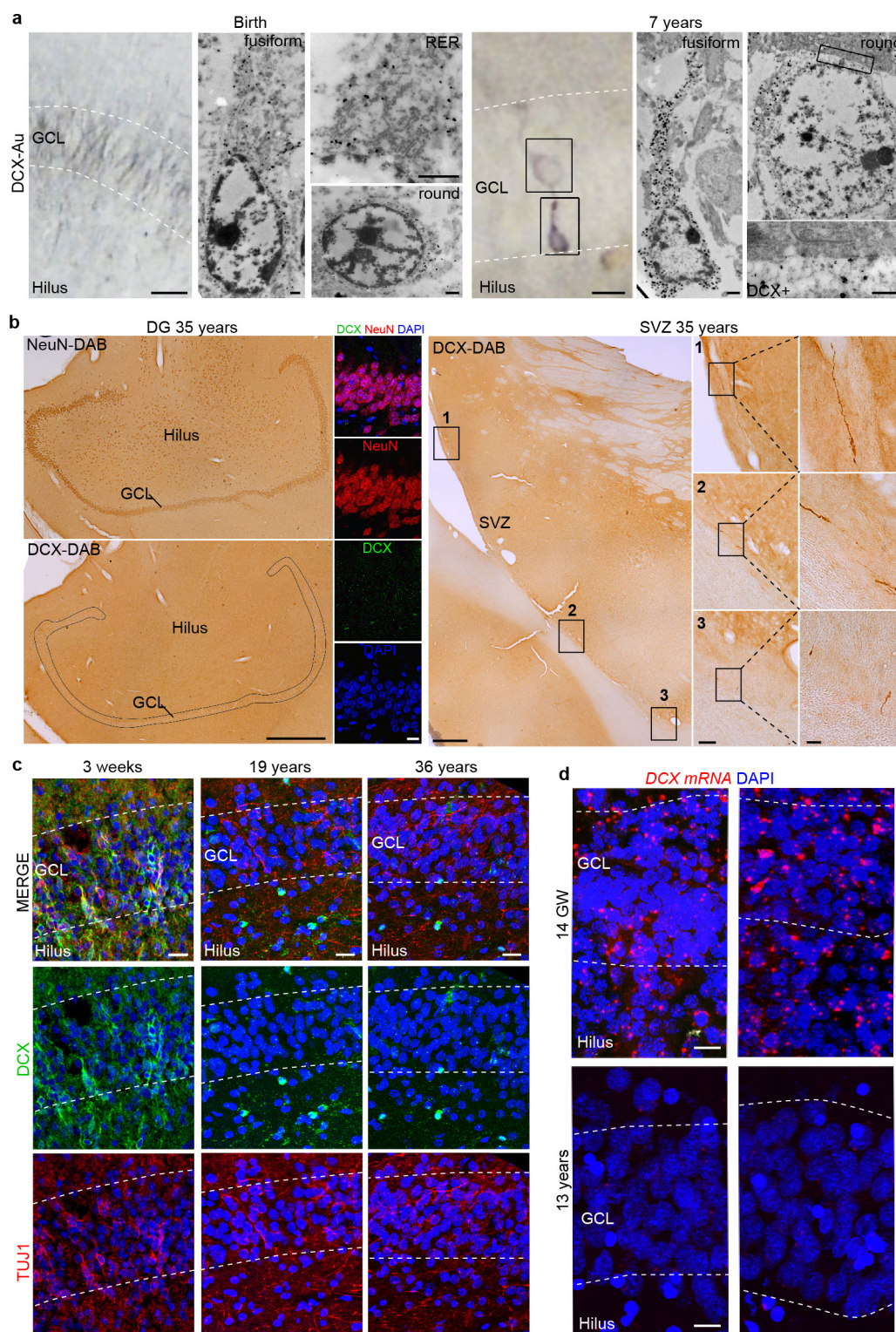
Extended Data Figure 4 | TEM analysis of cell types in the DG of human brains obtained from a 13-year-old individual and an adult; absence of SGZ precursor cells or immature neurons. **a**, Reconstruction of 5 ultrathin sections (separated by 1.5 μm) from the GCL of the 13-year-old individual with outlines of cell membranes. Colours corresponding to the different cell types defined by their ultrastructural characteristics are indicated in the key. No clusters or isolated cells with a young neuronal ultrastructure were found. Cells associated in small groups were identified as astrocytes, oligodendrocytes or microglia. **b**, **c**, Reconstructions of astroglial cells next to the GCL, searching for possible examples of radial astrocytes in the DG of an adult human. **b**, Example of an astrocyte with radial morphology in the adult GCL. Five serial semi-thin sections of

this astrocyte (black arrows) next to the GCL of the DG of a 48-year-old individual are shown; alternating semi-thin sections show that this cell is GFAP⁺. This cell extends a thin radial fibre through the GCL, but has multiple processes (stellate morphology) in the hilus. Boxed area shows the ultrastructure from the indicated semi-thin section of this astrocyte (pseudocoloured in blue) and the bundles of intermediate filaments present in the expansion (arrows). **c**, Another example of a serially reconstructed astrocyte in the DG of a 30-year-old individual with epilepsy (separated by 1.4 μm), showing a short radial expansion and processes into the hilus. Scale bars, 10 μm (**a**, **b**, semi-thin sections and TEM micrographs), 5 μm (**c**), 2 μm (**b**, soma) and 500 nm (**b**, intermediate filaments).



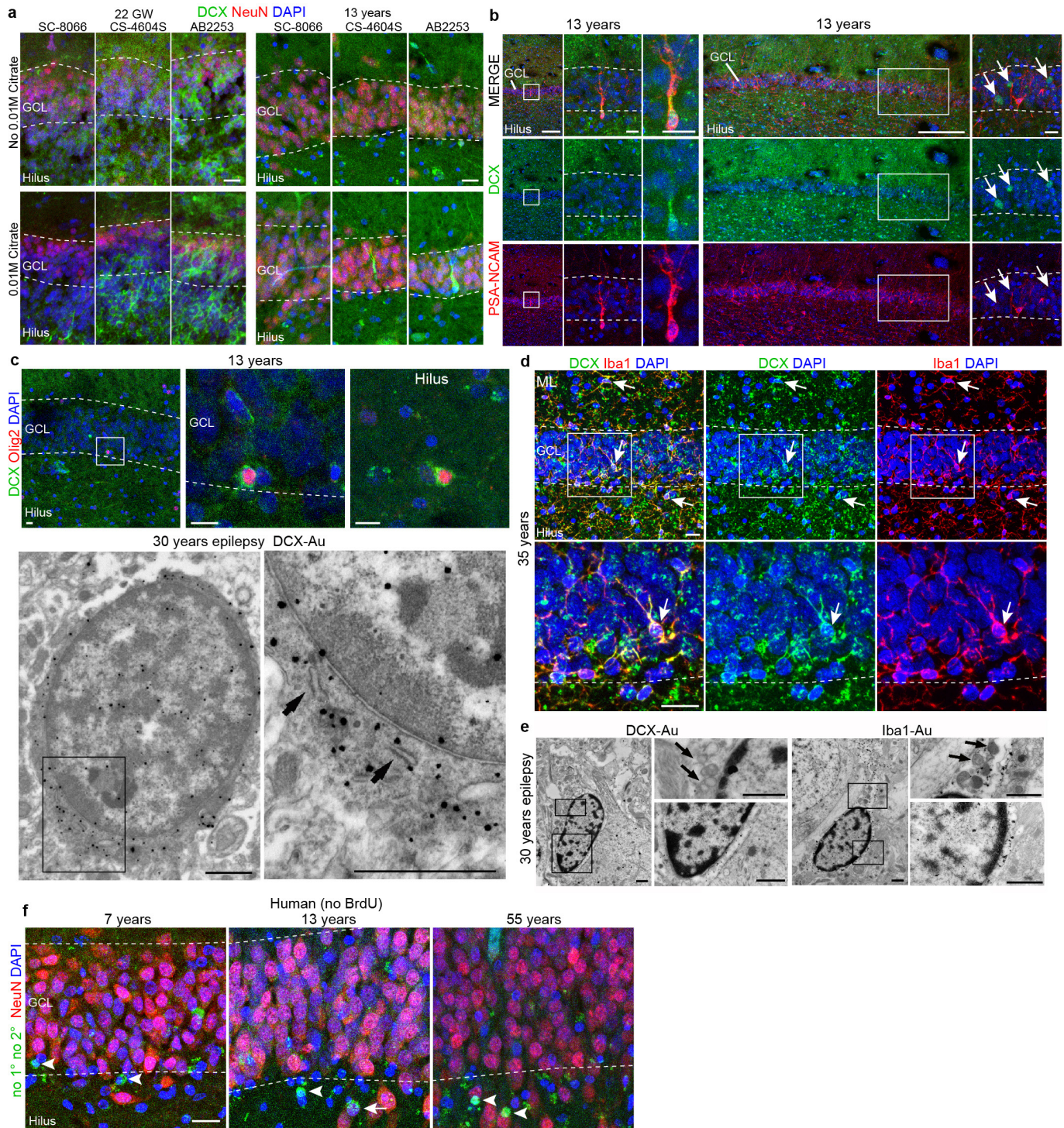
Extended Data Figure 5 | Young neurons are present in the infant but not the adult human DG. **a**, DAB staining in the hippocampus at birth reveals many young neurons in the GCL. **b**, DCX⁺PSA-NCAM⁺ cells are distributed in clusters across the GCL at 1 year of age. Most PSA-NCAM⁺ cells are DCX⁺, but some are DCX⁻PSA-NCAM⁺ (arrows). **c**, In the samples from a 13-year-old individual, DCX⁺ cells have a more mature neuronal morphology. The cell shown is NeuN⁺ and has dendrites in the molecular layer (arrowheads) and an axon projecting into the hilus (arrow). **d**, At 35 years of age, the DG does not contain DCX⁺PSA-NCAM⁺ cells, but does contain many DCX⁻PSA-NCAM⁺ cells that do

not have the morphology of young neurons. **e**, PSA-NCAM⁺ staining in the human DG from 3 weeks to 77 years; in adults, these cells have a more mature neuronal morphology and are localized in the hilus. **f**, PSA-NCAM⁺ cells in the DG are NeuN⁺ in samples of 19- and 77-year-old individuals. **g**, At 3 weeks of age, the GCL and hilus were filled with clusters of DCX⁺NeuroD⁺ cells, and many of the DCX⁻ GCL neurons were NeuroD⁺. At 35 years, no DCX⁺NeuroD⁺ cells were observed; antibody labelling for NeuroD was non-specific. Scale bars, 200 μ m (**a**, **d**–**g**), 20 μ m (**b**, **c**, and **d**, **f**, **g** (inset)).



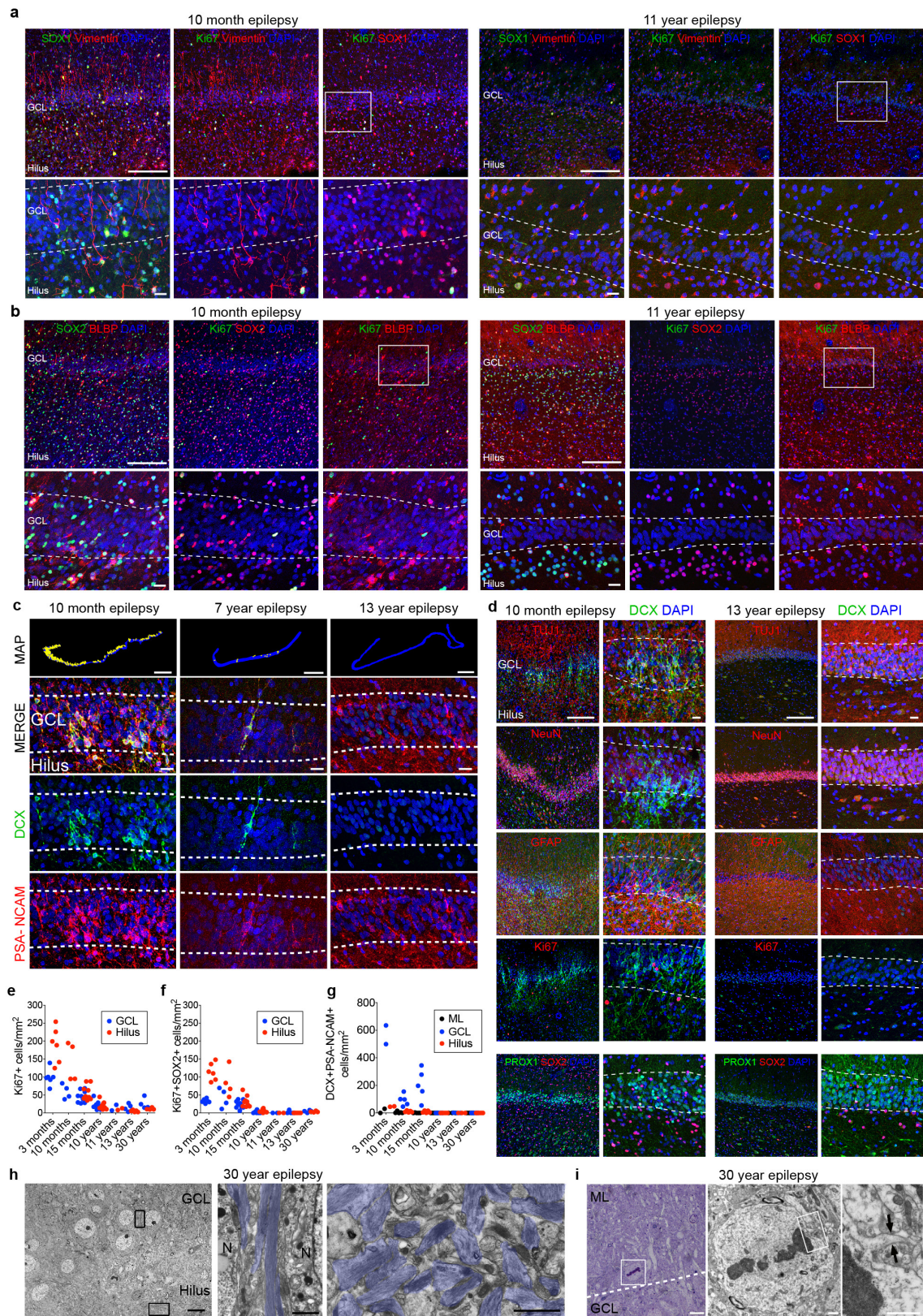
Extended Data Figure 6 | DCX⁺ young neurons in the developing human hippocampus. **a**, TEM micrographs of DCX immunogold staining at birth and 7 years of age. At birth, the GCL contains small DCX⁺ cells with little cytoplasm, rough endoplasmic reticulum (RER) cisternae and a fusiform or round nucleus. At 7 years of age, DCX⁺ cells closer to the hilus have characteristics of immature neurons, including few organelles and a long expansion towards the GCL. DCX⁺ cells located within the GCL have mature neuron characteristics, including a large, round nucleus, rough endoplasmic reticulum, mitochondria and microtubules consistent with a more mature neuronal morphology (see Extended Data Fig. 5). At higher magnification, the more mature-appearing DCX-labelled cells are adjacent to DCX⁻ GCL neurons. **b**, No DCX⁺ cells in the hilus and GCL (stained

by NeuN antibodies; left insets) were found in the brain of a 35-year-old individual that showed exceptional preservation. In this sample, rare DCX⁺ cells with the features of young migratory neurons were present in the ventricular-subventricular zone (right insets). SVZ, subventricular zone. **c**, DCX⁺TUJ1⁺ cells were present in the GCL and hilus at 3 weeks of age, but were not detected in the adult DG. **d**, RNA-scope detection of DCX mRNA revealed many cells in the DG at 14 gestational weeks, but weakly labelled cells distributed throughout in the DG and other regions of the hippocampus at 13 years of age. Scale bars: 1 mm (**b** (left)), 100 μ m (**b** (middle right inset)), 20 μ m (**b** (right insets), **c**), 10 μ m (**d**), 5 μ m (**a** (left)) and 500 nm (**a** (right, TEM)).



Extended Data Figure 7 | DCX⁺PSA-NCAM⁻ glial cells in the adult human hippocampus. **a**, Comparison of citrate antigen retrieval using three DCX antibodies from this study (SC-8066, CS-4604S and AB2253) in the GCL obtained from individuals at 22 gestational weeks and 13 years of age. The 13 year old DCX⁺ cell (Extended Data Fig. 5c) is shown in the lower right panel and adjacent sections were stained with the other antibodies. **b**, Example of a DCX⁺PSA-NCAM⁺ neuron in the GCL and DCX⁺PSA-NCAM⁻ staining in the sample from the 13-year-old individual (arrows). **c**, Examples of DCX⁺OLIG2⁺ cells in the GCL and hilus of the 13-year-old individual. Immunogold-labelled DCX⁺ cells viewed by TEM had single short endoplasmic reticulum cisternae (arrows), a very irregular contoured membrane and a round nucleus with condensed chromatin characteristic of oligodendrocytes. **d**, In some samples (see Extended Data Fig. 5g, bottom right inset), we found DCX⁺ immunoreactivity in many small multipolar cells. This staining

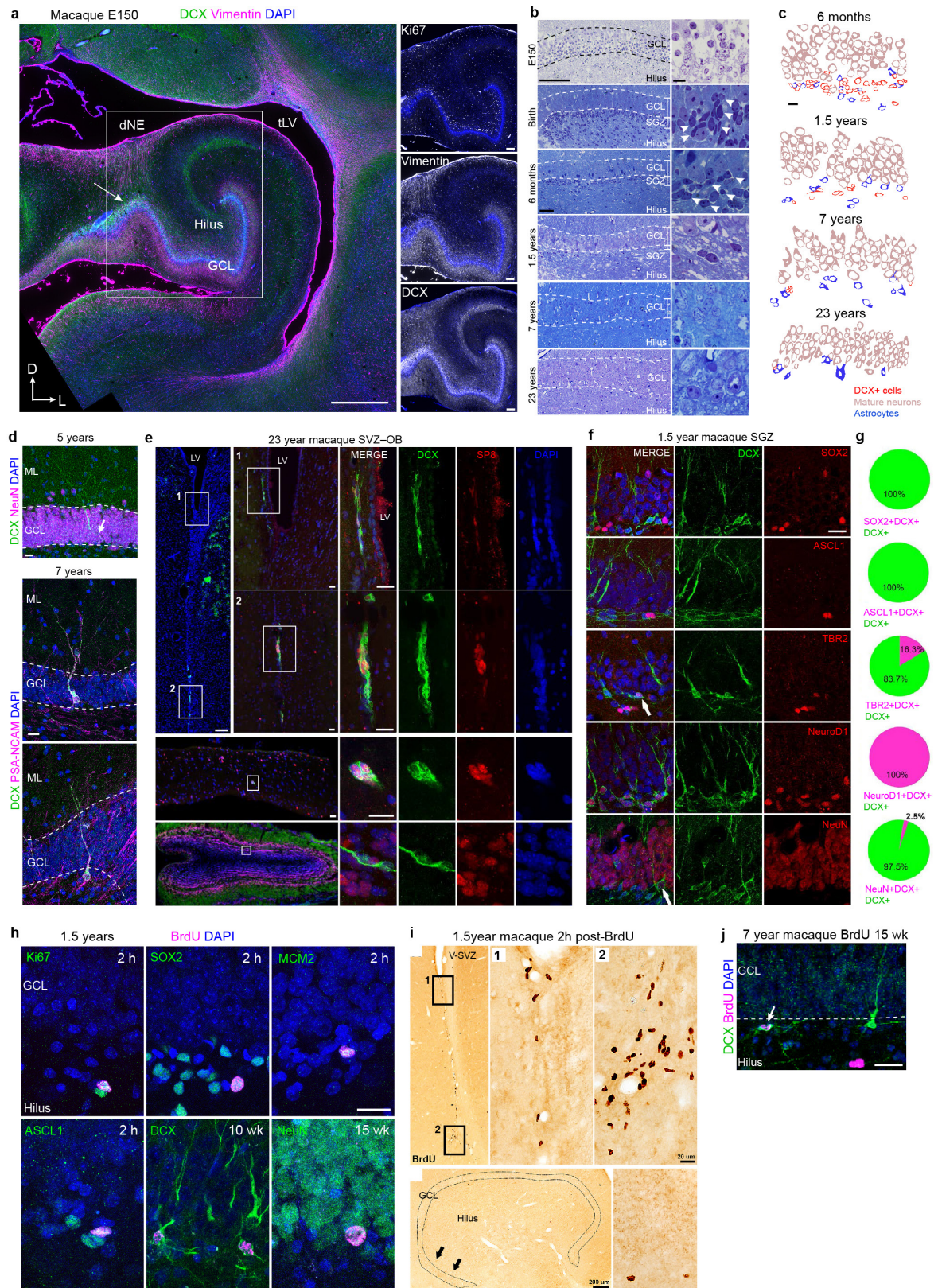
was not limited to the hilus, or GCL, but was present in cells across the tissue and co-localized with the microglial marker IBA1 (arrows). **e**, TEM micrographs of DCX and IBA1 immunogold-labelled cells in an adult DG of a 30-year-old individual with epilepsy. DCX⁺ and IBA1⁺ cells have similar characteristics: elongated nucleus with clumps of chromatin beneath the nuclear envelope and throughout the nucleoplasm, irregular contour and the presence of lysosomes and lipofuscin (arrows). Note that these features are typical of microglial cells. **f**, Human hippocampus stained with NeuN followed by processing for BrdU detection (with no primary or secondary antibodies) shows round fluorescent signal (arrowheads indicate signal that is NeuN⁻) occasionally overlapping with NeuN staining (arrow). Scale bars, 200 μ m (**b** (left column and wide column)), 20 μ m (**a**, **b** (left-middle columns, right column), **d**, **f**), 10 μ m (**c** (top row)) and 1 μ m (**c** (bottom row), **e**).



Extended Data Figure 8 | See next page for caption.

Extended Data Figure 8 | Neurogenesis declines in patients with epilepsy from infancy into childhood. **a**, Ki-67⁺SOX1⁺vimentin⁺ cells are located in the hilus and GCL at 10 months but are not present at 11 years of age. **b**, Ki-67⁺SOX2⁺BLBP⁺ cells are located in the hilus and GCL at 10 months but are not present at 11 years of age. **c**, Maps of DCX⁺PSA-NCAM⁺ cells (yellow dots) and representative immunostaining at 10 months, 7 years and 13 years (bottom rows). **d**, In the 10-month-old DG of a patient with epilepsy, DCX⁺ cells co-expressing PSA-NCAM or TUJ1 are distributed throughout the NeuN⁺PROX1⁺ GCL, but do not co-express Ki-67 or GFAP. In the DG of a 13-year-old patient with epilepsy, DCX⁺ cells co-expressing PSA-NCAM or TUJ1 were not present. Few Ki-67⁺ cells were visible throughout the DG. **e–g**, Quantification of Ki-67⁺ (**e**), Ki-67⁺SOX2⁺ (**f**) and DCX⁺PSA-NCAM⁺ (**g**) cells in the DG of surgically resected hippocampuses. **h**, TEM micrographs of the brain of a 30-year-old patient with epilepsy showing astroglial expansions with

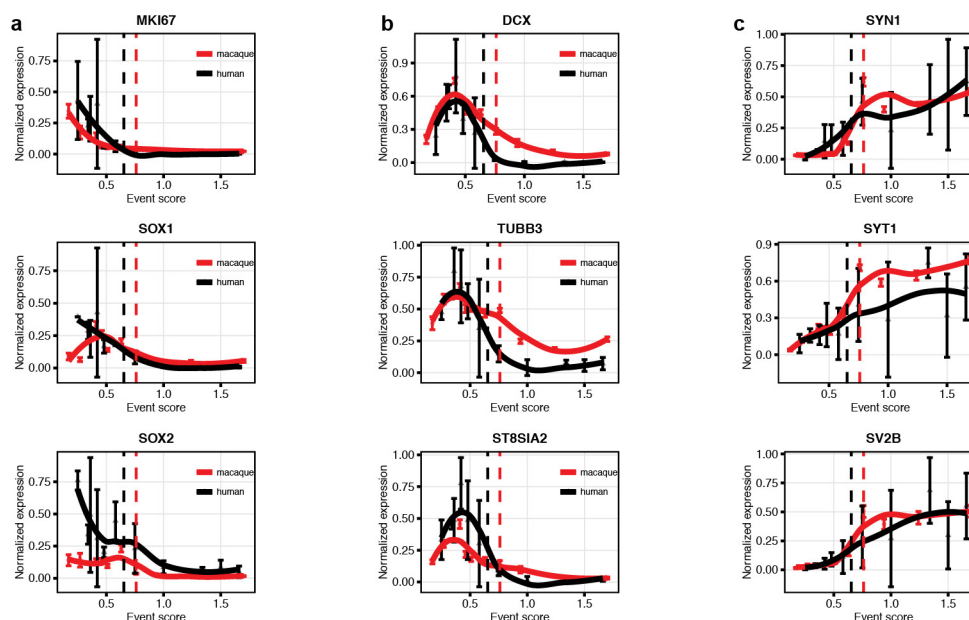
high number of intermediate filaments (blue) ensheathing GCL neuronal bodies. A dense network of astrocytic expansions in the hilus, containing dense bundles of intermediate filaments (blue), fills the region proximal to the GCL with no evidence of SGZ progenitor cells. **i**, Mitotic cells are very rare and not restricted to the hilus or GCL. A toluidine-blue-stained 1.5- μ m section from the DG of a 30-year-old brain shows a dividing cell in the molecular layer, adjacent to the GCL. The TEM micrograph shows the dividing cell in metaphase with a light cytoplasm, few organelles and an irregular contour with a small expansion (arrows), which are characteristic of astrocytes (shown at higher magnification). N, neuron. For quantifications, staining replicates (≥ 3) are shown by dots (each age, $n = 1$). Scale bars, 1 mm (**c** (maps)), 200 μ m (**a**, **b** (top), **d** (left)), 20 μ m (**a–c** (bottom), **d** (right)) and 10 μ m (**h**, **i** (left)), 1 μ m (**h**, **i** (middle and right)).



Extended Data Figure 9 | See next page for caption.

Extended Data Figure 9 | Development of the macaque DG and evidence of the presence of a proliferative SGZ and postnatal neurogenesis with a sharp decline in adulthood. **a**, The E150 macaque hippocampus has many Ki-67⁺ cells in the SGZ as well as vimentin⁺ fibres and DCX⁺ cells between the dNE and the GCL (arrow). **b**, Toluidine-blue-stained semi-thin sections of the macaque DG reveal small and dark condensed nuclei (arrows) in the SGZ from fetal ages (E150) to 1.5 years; few are visible in the DG at 7 years of age and these cells are very rare in the DG of a 23-year-old macaque (compare with human data in Extended Data Fig. 2a). **c**, Profiles of the cellular populations in the macaque DG at 6 months, 1.5, 7 and 23 years of age. As in the human DG, DCX⁺ cells decrease markedly with age and have little cytoplasm and a smaller nucleus compared to mature granule neurons. **d**, Top, example of a DCX⁺NeuN⁺ cell with mature dendrites in the GCL of a 5-year-old macaque. Middle, bottom, two examples of DCX⁺PSA-NCAM⁺ cells with dendritic arborization present in the GCL of a 7.5-year-old macaque. An axonal extension (arrow) into the hilus is visible. **e**, The ventricular–

subventricular zone (SVZ) and olfactory bulb (OB) of a 23-year-old macaque contain some DCX⁺SP8⁺ cells with the morphology of young neurons, but similar cells are rare in the GCL of the 23-year-old macaque (Fig. 4b, d). **f**, DCX⁺ cells in the SGZ of a 1.5-year-old macaque express transcription factors in common with those in the mouse SGZ. **g**, Percentages of DCX⁺ cells expressing markers shown in **f**. **h**, Immunostaining images of BrdU⁺ cells and cell proliferation (Ki-67, MCM2), progenitor cell markers (SOX2, ASCL1) or DCX, in the 1.5-year-old macaque euthanized 2 h after BrdU injection. BrdU⁺DCX⁺ and BrdU⁺NeuN⁺ cells could be identified 10 or 15 weeks after BrdU exposure, respectively. **i**, DAB-staining for BrdU in the ventricular–subventricular zone and SGZ of a 1.5-year-old macaque, 2 h after BrdU injection. **j**, Example of a rare DCX⁺BrdU⁺ cell in the 7.5-year-old macaque. Scale bars, 1 mm (**a** (left)) 200 μ m (**a** (right), **e** (top left, bottom left), **i** (left)), 100 μ m (**b** (left)), 20 μ m (**d**, **e** (middle left and right), **f**, **h**, **i** (right), **j**) and 10 μ m (**b** (right), **c**).



Extended Data Figure 10 | Decline in markers associated with neurogenesis in the macaque and human hippocampus (gene-expression profiling). **a**, Markers of dividing or precursor cells. **b**, Markers of young neurons. **c**, Markers of mature neurons. Human RNA-seq (<http://brainspan.org/>) and macaque expression profiling

(dataset from ref. 33) developmental data from hippocampus for the indicated genes. Human data are averaged over biological replicates by developmental period (as defined in ref. 34). Normalized data are plotted on the same developmental event scale. Loess-fit curves are displayed with data points (mean \pm s.e.m.). Dashed lines indicate birth.

Recognition of DHN–melanin by a C-type lectin receptor is required for immunity to *Aspergillus*

Mark H. T. Stappers^{1*}, Alexandra E. Clark^{1*}, Vishukumar Aimanian^{2*}, Stefan Bidula¹, Delyth M. Reid¹, Patawe Asamaphan¹, Sarah E. Hardison¹, Ivy M. Dambuza¹, Isabel Valsecchi², Bernhard Kerscher¹, Anthony Plato¹, Carol A. Wallace¹, Raif Yuecel³, Betty Hebecker¹, Maria da Glória Teixeira Sousa¹, Cristina Cunha^{4,5}, Yan Liu⁶, Ten Feizi⁶, Axel A. Brakhage⁷, Kyung J. Kwon–Chung⁸, Neil A. R. Gow¹, Matteo Zanda⁹, Monica Piras⁹, Chiara Zanato⁹, Martin Jaeger¹⁰, Mihai G. Netea¹⁰, Frank L. van de Veerdonk¹⁰, João F. Lacerda^{11,12}, António Campos Jr¹³, Agostinho Carvalho^{4,5}, Janet A. Willment¹, Jean–Paul Latgé^{2†} & Gordon D. Brown¹

Resistance to infection is critically dependent on the ability of pattern recognition receptors to recognize microbial invasion and induce protective immune responses. One such family of receptors are the C-type lectins, which are central to antifungal immunity¹. These receptors activate key effector mechanisms upon recognition of conserved fungal cell-wall carbohydrates. However, several other immunologically active fungal ligands have been described; these include melanin^{2,3}, for which the mechanism of recognition is hitherto undefined. Here we identify a C-type lectin receptor, melanin-sensing C-type lectin receptor (MelLec), that has an essential role in antifungal immunity through recognition of the naphthalene-diol unit of 1,8-dihydroxynaphthalene (DHN)–melanin. MelLec recognizes melanin in conidial spores of *Aspergillus fumigatus* as well as in other DHN-melanized fungi. MelLec is ubiquitously expressed by CD31⁺ endothelial cells in mice, and is also expressed by a sub-population of these cells that co-express epithelial cell adhesion molecule and are detected only in the lung and the liver. In mouse models, MelLec was required for protection against disseminated infection with *A. fumigatus*. In humans, MelLec is also expressed by myeloid cells, and we identified a single nucleotide polymorphism of this receptor that negatively affected myeloid inflammatory responses and significantly increased the susceptibility of stem-cell transplant recipients to disseminated *Aspergillus* infections. MelLec therefore recognizes an immunologically active component commonly found on fungi and has an essential role in protective antifungal immunity in both mice and humans.

C-type lectin receptors (CLRs) involved in antifungal immunity belong primarily to the Dectin-1 (also known as CLEC7A) and Dectin-2 (also known as CLEC4N) clusters of receptors located near the natural killer gene complex¹. To identify new CLRs within these clusters that recognize fungi, we generated soluble protein chimaeras consisting of the C-type lectin-like domain of murine receptors fused to the Fc region of human immunoglobulin G1⁴ (Fc-MelLec, Extended Data Fig. 1a) and used these chimaeric proteins as probes to screen for the recognition of fungi by flow cytometry. Using this approach, we identified the C-type lectin-like domain of MelLec (CLEC1A (ref. 5), Extended Data Fig. 1b), which bound *A. fumigatus* conidia (Fig. 1a). MelLec did not recognize other commonly occurring fungi, such as

Candida albicans yeast and filamentous cells, or *Saccharomyces cerevisiae* yeasts, but did recognize other melanized fungal species, including *Fonsecaea pedrosoi* and *Cladosporium cladosporioides* (Fig. 1b, Extended Data Fig. 1c, e). This indicated that the ligand recognized by MelLec was not ubiquitously found in all fungi, unlike the ligands of other antifungal CLRs such as Dectin-1 (ref. 1). Notably, the ability of MelLec to recognize *A. fumigatus* was restricted to conidia, and recognition was rapidly lost after conidial swelling, germination and hyphal growth (Fig. 1a, c and Extended Data Fig. 2d). The binding to conidia was visualized by immunofluorescence microscopy, which revealed a punctate staining pattern suggestive of a restricted distribution of the moiety recognized by MelLec on these spores (Fig. 1c). Conidia are covered by a hydrophobic rodlet layer, which masks underlying components of the cell wall from immune recognition⁶. The removal of this rodlet layer with sodium hydroxide⁶ led to increased and uniform binding of MelLec over the entire conidial surface (Fig. 1d). Indeed, uniform staining was obtained with rodlet-deficient conidia ($\Delta rodA$), confirming that the ligand of MelLec is partially masked by the surface hydrophobin layer (Fig. 1e). Recognition of *A. fumigatus* conidia by MelLec was also demonstrated in a cellular context using MelLec reporter cells⁴ (Extended Data Fig. 1f).

All characterized CLRs that are involved in fungal sensing act by recognition of carbohydrate components of the fungal cell wall¹. Consistent with this possibility, MelLec ligand(s) were detected by enzyme-linked immunosorbent assay (ELISA) primarily in the alkali-insoluble fraction of the *A. fumigatus* conidial cell wall, the primary constituents of which are carbohydrates (β -glucan, chitin and galactomannan) and melanin⁷ (Fig. 2a). The Fc-MelLec fusion protein was used as a probe to screen a neoglycolipid-based glycan microarray containing almost 500 structures, including oligosaccharides derived from glucans and chitin that are found in fungi⁸ (Supplementary Table 1). This screen did not reveal any carbohydrate ligands for MelLec; ligands were detected, however, for the C-type lectin langerin, which was used as a control (Extended Data Fig. 2). We then used Fc-MelLec to screen *A. fumigatus* conidia that were mutated in various relevant cell-wall biosynthetic pathways, using flow cytometry and immunofluorescence microscopy. This revealed that Fc-MelLec failed to detect $\Delta pksP$ mutant conidia, which are deficient in the ability to synthesize heptaketide naphthopyrone (YWA1), the first intermediate of the

¹Medical Research Council Centre for Medical Mycology at the University of Aberdeen, Aberdeen Fungal Group, Institute of Medical Sciences, Foresterhill, Aberdeen AB25 2ZD, UK. ²Unité des Aspergillus, Institut Pasteur, Paris, France. ³Inain Fraser Cytometry Centre, Institute of Medical Sciences, University of Aberdeen, Aberdeen AB25 2ZD, UK. ⁴Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, Braga, Portugal. ⁵ICVS/3B's – PT Government Associate Laboratory, Braga/Guimarães, Portugal. ⁶Glycosciences Laboratory, Department of Medicine, Imperial College London, London W12 0NN, UK. ⁷Department of Microbiology and Molecular Biology, Leibniz Institute for Natural Product Research and Infection Biology (HKL), Friedrich Schiller University, D-07745 Jena, Germany. ⁸Molecular Microbiology Section, Laboratory of Clinical Immunology and Microbiology, National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH), Bethesda, Maryland, USA. ⁹Institute of Medical Sciences, University of Aberdeen, Foresterhill, Aberdeen AB25 2ZD, UK. ¹⁰Department of Internal Medicine, Radboud University Medical Center, Nijmegen, The Netherlands. ¹¹Instituto de Medicina Molecular, Faculdade de Medicina de Lisboa, Universidade de Lisboa, Lisboa, Portugal. ¹²Serviço de Hematologia e Transplantação de Medula, Hospital de Santa Maria, Lisboa, Portugal. ¹³Serviço de Transplantação de Medula Óssea (STMO), Instituto Português de Oncologia do Porto, Porto, Portugal. [†]Present address: State Key Laboratory, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China.

*These authors contributed equally to this work.

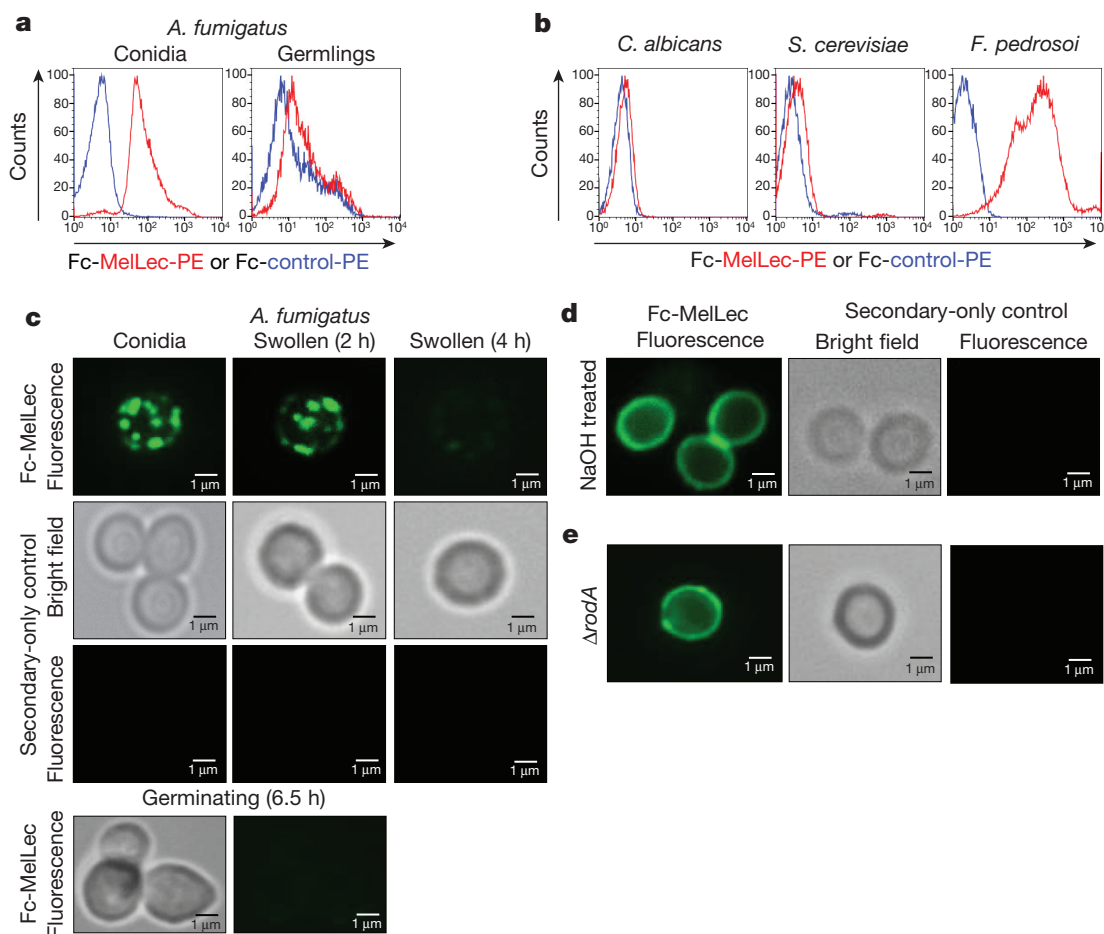


Figure 1 | MelLec recognizes selected fungi. **a, b**, Representative histograms showing *A. fumigatus* conidia and germlings (cultured for 8 h at 37 °C) (**a**) and yeasts of *C. albicans* and *S. cerevisiae*, and conidia of *F. pedrosoi* (**b**) stained with Fc-MelLec or Fc-CLEC12B (ref. 24) (Fc-control) and analysed by flow cytometry. **c–e**, Representative light microscopy

images and immunofluorescence micrographs using Fc-MelLec to detect ligands on *A. fumigatus*, after conidial swelling and germination over time (**c**), after treatment with 1 M NaOH (**d**) and on rodlet-deficient ($\Delta rodA$) conidia (**e**). Experiments were repeated at least three times independently, with similar results.

DHN-melanin biosynthetic pathway^{2,9} (Fig. 2b and Extended Data Fig. 3a). We also confirmed loss of the MelLec ligand on $\Delta rodA \Delta pksP$ double mutant conidia¹⁰—which lack both melanin and the hydrophobin layer—using flow cytometry, immunofluorescence microscopy and our MelLec reporter cells (Extended Data Figs 1f and 3b, c). Loss of *pksP* did not affect conidial recognition by Fc-Dectin-1 (Extended Data Fig. 3c, d). Moreover, we demonstrated a direct interaction of MelLec with *A. fumigatus* melanin ghosts (Extended Data Fig. 3e). MelLec recognized specifically DHN-melanin, as this receptor did not detect melanin synthesized by other pathways, such as those used by *Cryptococcus neoformans*, and mammalian B16 melanoma cells¹¹ (Extended Data Fig. 3f). In addition to *A. fumigatus*, we showed that MelLec recognized other pigmented fungi that produce DHN-melanin, including *F. pedrosoi* and *C. cladosporioides*¹² (Fig. 1b and Extended Data Fig. 1e).

Although its tertiary structure is unresolved¹¹, the biosynthetic pathway of DHN-melanin is well characterized in *A. fumigatus*³. To determine at which stage MelLec ligand(s) are synthesized in the pathway, we screened mutants of *A. fumigatus* that were deficient in the enzymes required to catalyse each step, using Fc-MelLec and immunofluorescence microscopy, or ELISA. We confirmed that MelLec recognition was lost in $\Delta pksP$ conidia; however, MelLec recognition was unaffected in mutants deficient in the enzymes required for all other stages of the DHN-melanin biosynthetic pathway (Fig. 2c and Extended Data Fig. 4a). Using a pre-adsorption assay, we showed that the recognition of wild-type conidia could be inhibited by pretreating Fc-MelLec with $\Delta ayg1$ ghosts (defective in the second biosynthetic

step, in which MelLec ligands are still present), but not ghosts of $\Delta pksP$ (the first biosynthetic step, which lacks MelLec ligands; Extended Data Fig. 4b). We purified YWA1 from $\Delta ayg1$ conidia¹³ and demonstrated that pretreatment with this compound inhibited the ability of Fc-MelLec, but not the ability of Fc-Dectin-1, to detect $\Delta rodA$ conidia or sodium hydroxide-treated wild-type conidia (Extended Data Fig. 4c, d). Moreover, using ELISA, we demonstrated a direct interaction of MelLec with purified YWA1 (Fig. 2d). This suggested that the ability of MelLec to recognize YWA1, as well as all of the other melanin biosynthetic intermediates (Fig. 2c), was due to recognition of the conserved naphthalene-diol unit, present in each of the intermediates. Indeed, we revealed a direct interaction of MelLec with 1,8-dihydronaphthalene (1,8-DHN), another upstream intermediate of the melanin biosynthetic pathway, by ELISA (Extended Data Fig. 4e). The structural isomers 1,2-DHN and 1,4-DHN, which contain the naphthalene-diol unit but are not melanin intermediates, were also recognized by MelLec (Extended Data Fig. 4f). This suggests that the position of at least one of the hydroxyls (on carbon 2, 4 or 8) on the naphthalene-diol unit is not important for recognition by MelLec. By contrast, naphthalene and 1-naphthol were not recognized by MelLec (Extended Data Fig. 4g), indicating that the ligand of MelLec is a naphthalene-diol.

CLRs involved in fungal recognition are predominantly expressed by myeloid cells and are detectable in most tissues¹. Using reverse transcription PCR (RT-PCR), we found that MelLec was widely expressed in mice, with the highest levels of transcript detected in the lung (Extended Data Fig. 5a). To explore expression at a cellular level,

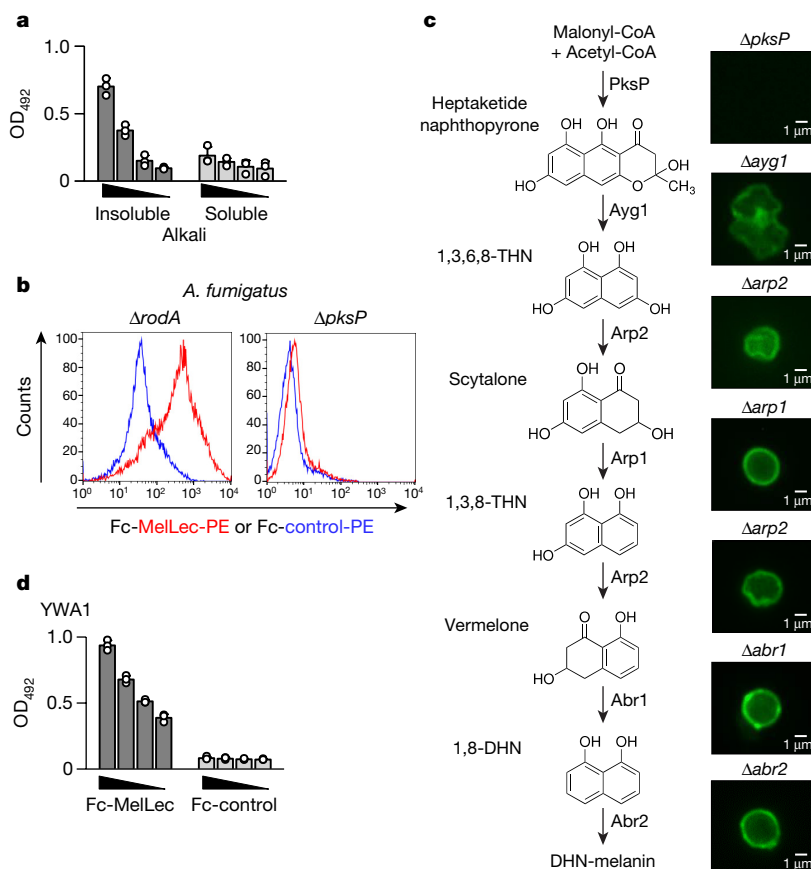


Figure 2 | MelLec recognizes DHN-melanin. **a**, Detection of MelLec ligands in alkali-insoluble and alkali-soluble *A. fumigatus* cell-wall fractions by ELISA. Values show mean \pm s.d. **b**, Representative histograms showing $\Delta rodA$ or $\Delta pksP$ *A. fumigatus* conidia, stained with Fc-MelLec or Fc-CLEC12B (Fc-control) and analysed by flow cytometry. **c**, The biosynthetic pathway of DHN-melanin (left), and representative

immunofluorescence micrographs obtained using Fc-MelLec to detect ligands on *A. fumigatus* strains deficient in the stated enzymes (right). The conidial rodlet layer was removed with 1 M NaOH before staining. **d**, Detection of YWA1 by Fc-MelLec and Fc-control by ELISA. Values show mean \pm s.d. Experiments were repeated at least three times independently, with similar results.

we generated monoclonal antibodies by immunizing rats with murine Fc-MelLec and then screening ELISA-positive hybridoma supernatants using flow cytometry on NIH3T3 cells transfected to express full-length haemagglutinin-tagged murine MelLec. In these transfected cells, we found that MelLec was expressed as a glycosylated monomer at the cell surface, which demonstrated that this CLR did not require an adaptor for surface expression^{1,14} (Extended Data Fig. 5b, c). Although MelLec is able to sense melanin (Fig. 2 and Extended Data Fig. 1f), the expression of this receptor on these transfected NIH3T3 cells did not confer the ability to capture *A. fumigatus* conidia to the cell surface (Extended Data Fig. 5d).

Two monoclonal antibodies (18E4 and 14C8) specific for MelLec were chosen for further characterization of receptor expression on mouse cells and tissues (Extended Data Fig. 5e). Surprisingly, MelLec was not expressed by any examined mouse myeloid cell population (either *ex vivo* or *in vitro* bone marrow-derived), even after microbial stimulation, nor was this receptor expressed by cells in peripheral blood, bone marrow, lymphoid tissues or on platelets¹⁵ (Extended Data Fig. 6a–c). Given the abundance of transcript, next we examined disaggregated lung tissue by flow cytometry, which revealed a distinct population of cells that expressed MelLec (Fig. 3a). Histological visualization of MelLec expression by immunofluorescence microscopy revealed broad punctate expression of this receptor throughout the lung tissue (Fig. 3b and Extended Data Fig. 7m). A similar punctate staining pattern was also observed in the MelLec-expressing transfected NIH3T3 fibroblasts (Extended Data Fig. 7a). Flow cytometry analysis showed that MelLec expression was restricted to non-haematopoietic (CD45[−]) cells (Fig. 3c and Extended Data Fig. 7b). Further

characterization of these cells revealed that MelLec was expressed by CD31⁺ EpCAM[−] endothelial cells (EpCAM, epithelial cell adhesion molecule) (Fig. 3d). Expression of MelLec was also detected on CD31⁺ endothelial cells in all other tissues tested, including the liver, heart, kidney and small intestine (Extended Data Fig. 7c–f). However, we also detected a unique population of CD31⁺ cells that co-expressed EpCAM⁺, but only in the lung and the liver, and these cells also expressed MelLec (Fig. 3d and Extended Data Fig. 7c). MelLec was not expressed on EpCAM⁺ cells in other tissues, including in the epidermis (Extended Data Fig. 7g–j).

To gain insight into the physiological functions of MelLec, we generated mice deficient in this receptor using a conventional gene-targeting vector (Extended Data Fig. 7k, l). Exons 1–5 of the gene encoding MelLec (*Clec1a*) were deleted; these exons correspond to the cytoplasmic tail, transmembrane, stalk and part of the CRD region. Flow cytometry of disaggregated lung tissue, and immunofluorescence microscopy of whole lung, confirmed the lack of MelLec expression in cells from knockout mice (Fig. 3b and Extended Data Fig. 7m, n). The MelLec-knockout mice were viable, had no gross abnormalities and had normal peripheral leucocyte counts (Supplementary Table 2). Intratracheal (i.t.) challenge of immunocompetent MelLec-knockout mice with wild type *A. fumigatus* conidia revealed no alterations in survival or in other physiological parameters, including weight (Extended Data Fig. 8a). However, a significantly reduced influx of neutrophils into the lungs of MelLec-knockout mice could be detected shortly (4 h) after i.t. challenge (Extended Data Fig. 8b–d), before the conidia had germinated¹⁶. This alteration in cellular recruitment in the MelLec-knockout mice was associated with alterations in selected

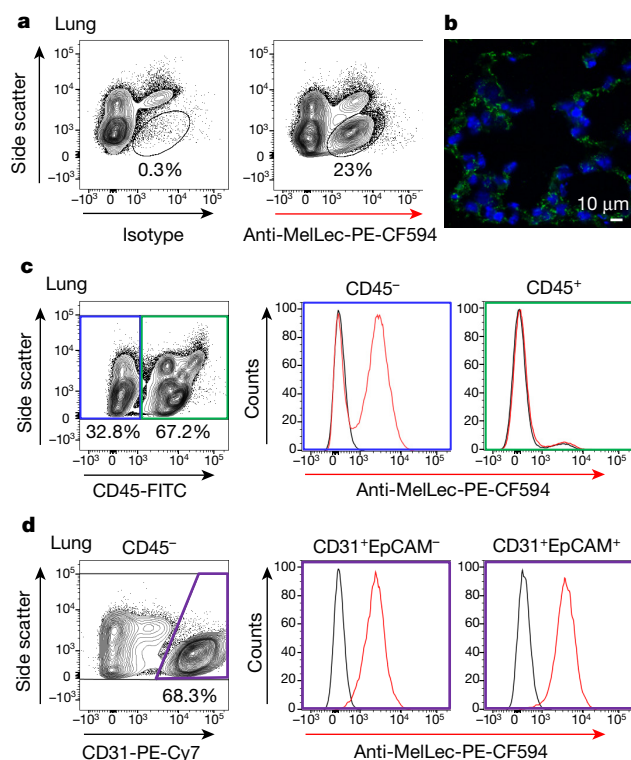


Figure 3 | MelLec is expressed on non-myeloid cells in mouse.

a, Analysis of disaggregated lung tissue by flow cytometry with anti-MelLec antibody. **b**, Immunofluorescence micrographs of lung tissue stained with anti-MelLec antibody (green). Nuclei are stained with 4',6-diamidino-2-phenylindole (DAPI) (blue). **c**, **d**, Flow-cytometric analysis of MelLec expression on live CD45⁺ and CD45⁻ cells (**c**), and CD45⁻ CD31⁺EpCAM⁻, and CD45⁻ CD31⁺EpCAM⁺ cells (**d**) in the lung. Experiments were repeated at least three times independently, with similar results.

neutrophil-related cytokines, including KC (CXCL1) and granulocyte macrophage colony-stimulating factor (Extended Data Fig. 8e). A similar alteration in neutrophil recruitment was also observed in the MelLec-knockout mice after i.t. challenge with melanin ghosts of *A. fumigatus* (Extended Data Fig. 8f). There were no alterations in other pulmonary myeloid populations after conidial challenge, and by 24 h after the challenge the difference in neutrophil influx was no longer apparent (Extended Data Fig. 8g, h). There was also no change in the expression of MelLec during infection with *A. fumigatus* (Extended Data Fig. 8i). Notably, the defect in early neutrophil-recruitment in the MelLec-knockout mice was lost upon i.t. challenge with Δ *pksP* conidia, which lack melanin (Extended Data Fig. 8j, k).

We next examined the role of MelLec during infections in corticosteroid-treated mice, to model the effects of immunosuppression¹⁷. Under these conditions, loss of MelLec did not significantly alter the susceptibility to infection (Extended Data Fig. 9a). However, when we intravenously (i.v.) infected immunocompetent MelLec-knockout mice with *A. fumigatus* conidia¹⁷, we observed substantially increased susceptibility in these mice compared to the wild type (Fig. 4a). This increased susceptibility was associated with increased fungal burdens in several tissues, including the brain (Fig. 4b and Extended Data Fig. 9b–d), as well as alterations in inflammatory responses (Fig. 4c). IL-17 responses were unaffected¹⁸. Consistent with a role in melanin recognition, there was no difference in susceptibility or fungal burden between wild-type and MelLec-knockout mice after systemic infection with Δ *pksP* conidia (Fig. 4d and Extended Data Fig. 9e).

Our mouse data suggested that MelLec has a key role in the immunity to disseminated infections with *Aspergillus*. Therefore, we next explored the role of this receptor in humans. MelLec has been

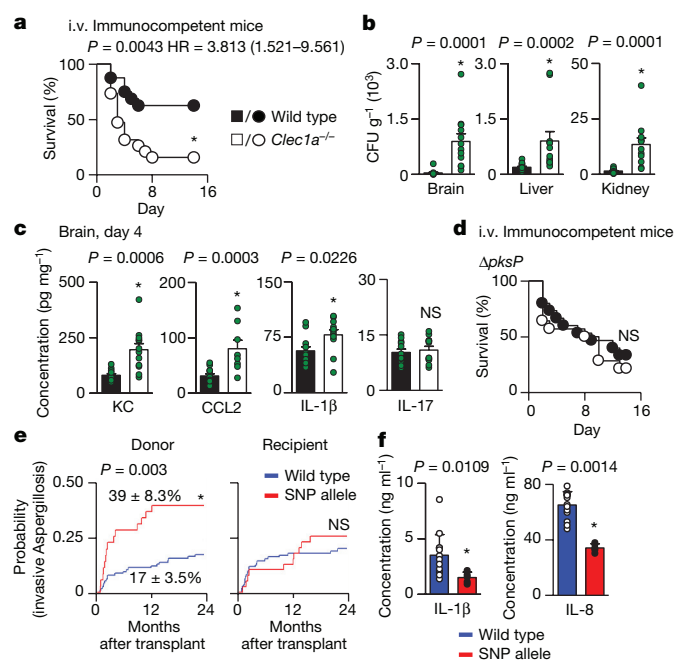


Figure 4 | MelLec is required to prevent disseminated infection in mice and humans. **a**, Survival of mice after i.v. infection with 10^6 *A. fumigatus* conidia ($n = 16$ mice per group). Pooled data from two independent experiments, analysed by log-rank test. **b**, **c**, Tissue fungal burdens (**b**) and brain cytokine levels (**c**) of mice four days after i.v. infection with 10^6 *A. fumigatus* conidia ($n = 12$ mice per group). CFU, colony-forming unit. Values shown are mean \pm s.e.m. of pooled data from two independent experiments, analysed by two-sided Mann–Whitney *U* test. Green circles indicate individual data points in all cases. **d**, Survival of mice after i.v. infection with 10^6 Δ *pksP* *A. fumigatus* conidia ($n = 15$ mice per group). Pooled data from two independent experiments, analysed by log-rank test. **e**, Cumulative incidence analysis of invasive aspergillosis after transplantation according to donor (wild-type, $n = 238$ individuals; SNP allele, $n = 72$ individuals) or recipient (wild-type, $n = 228$ individuals; SNP allele, $n = 80$ individuals) *CLEC1A* rs2306894 genotypes, analysed by two-sided Gray's test. **f**, Cytokine production in monocyte-derived macrophages, after stimulation with *A. fumigatus* conidia (wild-type, $n = 14$ individuals; SNP allele, $n = 5$ individuals). Values shown are mean \pm s.d., analysed by two-sided Mann–Whitney *U* test. * $P \leq 0.05$; NS, not significant; HR, hazard ratio.

previously detected in humans on endothelial cells^{19,20} but also on myeloid cells^{5,18–20}, and we found that a human MelLec Fc fusion protein (Fc-hMelLec) recognizes DHN-melanized conidia (Extended Data Fig. 10a). A common missense single nucleotide polymorphism (SNP) within the coding region of *CLEC1A* (rs2306894, global minor allele frequency = 0.3295) results in an amino acid change (Gly26Ala) in the cytoplasmic tail of human MelLec. Notably, we found a highly significant association between this SNP and the risk of aspergillosis in stem-cell transplant recipients (Fig. 4e). This increased risk occurred when the variant was carried by the donor, but not when it was carried by the recipient. This suggests that, in humans, the protective functions of MelLec are primarily mediated by myeloid cells. In our mouse model, there was no difference in resistance to infection upon adoptive transfer of MelLec-deficient bone marrow into irradiated wild-type recipients (Extended Data Fig. 10b).

To demonstrate the effect of the SNP in *CLEC1A* and subsequent amino acid substitution in MelLec on the function of myeloid cells, we analysed the responses of monocyte-derived macrophages isolated from healthy genotyped donors. We found that macrophages from the individuals carrying this SNP produced significantly less IL-1 β and IL-8 after *in vitro* stimulation with *A. fumigatus* conidia compared to controls (Fig. 4f), whereas there was no difference in response upon stimulation with lipopolysaccharide (Extended Data Fig. 10c). We verified

the inflammatory defect caused by this SNP in peripheral blood mononuclear cells isolated from an independent cohort (Extended Data Fig. 10d). Moreover, using transduced RAW264.7 macrophages, we demonstrated directly that this SNP results in an inflammatory defect upon stimulation with melanin-containing conidia (Extended Data Fig. 10e).

Melanin is considered a fungal virulence factor, providing protection against reactive oxygen species and inhibiting host-cell phagocytosis, cytokine production and apoptosis^{3,10,21}. Here we show that fungal DHN-melanin is also sensed by the host, through a melanin-sensing C-type lectin receptor (MelLec) that has a crucial role in the control of systemic *A. fumigatus* infection in both mice and humans. However, the data presented here, as well as early studies in rats^{18,22}, show that the cellular expression of this receptor differs between species. Critically, we define a polymorphism of this receptor that, when present in donor cells, increases the susceptibility of stem-cell transplant recipients to disseminated aspergillosis. Our data therefore suggest that identifying donors carrying this SNP could help considerably to reduce the incidence of this disease in transplant recipients. It is likely that MelLec will have an important role in immunity to other melanized fungi and black yeasts²³, especially those that cause phaeohyphomycosis, mycetoma and chromoblastomycosis.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 3 October 2016; accepted 6 February 2018.

Published online 28 February 2018.

- Hardison, S. E. & Brown, G. D. C-type lectin receptors orchestrate antifungal immunity. *Nat. Immunol.* **13**, 817–822 (2012).
- Nosanchuk, J. D. & Casadevall, A. The contribution of melanin to microbial pathogenesis. *Cell. Microbiol.* **5**, 203–223 (2003).
- Heinekamp, T. et al. *Aspergillus fumigatus* melanins: interference with the host endocytosis pathway and impact on virulence. *Front. Microbiol.* **3**, 440 (2013).
- Pyž, E. & Brown, G. D. Screening for ligands of C-type lectin-like receptors. *Methods Mol. Biol.* **748**, 1–19 (2011).
- Colonna, M., Samaridis, J. & Angman, L. Molecular characterization of two novel C-type lectin-like receptors, one of which is selectively expressed in human dendritic cells. *Eur. J. Immunol.* **30**, 697–704 (2000).
- Aimanianda, V. et al. Surface hydrophobin prevents immune recognition of airborne fungal spores. *Nature* **460**, 1117–1121 (2009).
- Latgé, J. P. & Beauvais, A. Functional duality of the cell wall. *Curr. Opin. Microbiol.* **20**, 111–117 (2014).
- Palma, A. S. et al. Ligands for the β -glucan receptor, Dectin-1, assigned using “designer” microarrays of oligosaccharide probes (neoglycolipids) generated from glucan polysaccharides. *J. Biol. Chem.* **281**, 5771–5779 (2006).
- Jahn, B. et al. Isolation and characterization of a pigmentless-conidium mutant of *Aspergillus fumigatus* with altered conidial surface and reduced virulence. *Infect. Immun.* **65**, 5110–5117 (1997).
- Akoumianaki, T. et al. *Aspergillus* cell wall melanin blocks LC3-associated phagocytosis to promote pathogenicity. *Cell Host Microbe* **19**, 79–90 (2016).
- Nosanchuk, J. D., Stark, R. E. & Casadevall, A. Fungal melanin: what do we know about structure? *Front. Microbiol.* **6**, 1463 (2015).
- Llorente, C. et al. *Cladosporium cladosporioides* LPSC 1088 produces the 1,8-dihydroxynaphthalene-melanin-like compound and carries a putative *pks* gene. *Mycopathologia* **174**, 397–408 (2012).
- Tsai, H. F. et al. Pentaketide melanin biosynthesis in *Aspergillus fumigatus* requires chain-length shortening of a heptaketide precursor. *J. Biol. Chem.* **276**, 29292–29298 (2001).
- Sancho, D. & Reis e Sousa, C. Signaling by myeloid C-type lectin receptors in immunity and homeostasis. *Annu. Rev. Immunol.* **30**, 491–529 (2012).
- Rambach, G. et al. Identification of *Aspergillus fumigatus* surface components that mediate interaction of conidia and hyphae with human platelets. *J. Infect. Dis.* **212**, 1140–1149 (2015).
- Faro-Trindade, I. et al. Characterisation of innate fungal recognition in the lung. *PLoS ONE* **7**, e35675 (2012).
- Clemons, K. V. & Stevens, D. A. The contribution of animal models of aspergillosis to understanding pathogenesis, therapy and virulence. *Med. Mycol.* **43** (Suppl. 1), 101–110 (2005).
- Lopez Robles, M. D. et al. Cell-surface C-type lectin-like receptor CLEC-1 dampens dendritic cell activation and downstream Th17 responses. *Blood Adv.* **1**, 557–568 (2017).
- Sobanov, Y. et al. A novel cluster of lectin-like receptor genes expressed in monocytic, dendritic and endothelial cells maps close to the NK receptor genes in the human NK gene complex. *Eur. J. Immunol.* **31**, 3493–3503 (2001).
- Sattler, S. et al. The human C-type lectin-like receptor CLEC-1 is upregulated by TGF- β and primarily localized in the endoplasmic membrane compartment. *Scand. J. Immunol.* **75**, 282–292 (2012).
- Chai, L. Y. et al. *Aspergillus fumigatus* conidial melanin modulates host cytokine response. *Immunobiology* **215**, 915–920 (2010).
- Thebault, P. et al. The C-type lectin-like receptor CLEC-1, expressed by myeloid cells and endothelial cells, is up-regulated by immunoregulatory mediators and moderates T cell activation. *J. Immunol.* **183**, 3099–3108 (2009).
- Seyedmousavi, S. et al. Black yeasts and their filamentous relatives: principles of pathogenesis and host defense. *Clin. Microbiol. Rev.* **27**, 527–542 (2014).
- Hoffmann, S. C. et al. Identification of CLEC12B, an inhibitory receptor on myeloid cells. *J. Biol. Chem.* **282**, 22370–22375 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the staff of the University of Aberdeen animal facility for their support and care for our animals, C. G. Park for providing recombinant langerin, and S. Filler and R. Cramer for advice. Funding was provided by the Wellcome Trust (102705, 097377, 093378, 099197, 108430, 101873), the Medical Research Council Centre for Medical Mycology and the University of Aberdeen (MR/N006364/1). K.J.K.-C. is supported by the intramural program of the National Institute of Allergy and Infectious Diseases, National Institutes of Health; V.A. by an ANR-DST COMASPIN grant (ANR-13-ISO3-0004); B.H. by German Science Foundation (www.dfg.de) grant no. HE 7565/1-1; J.-P.L., I.V. and V.A. by the ANR and FRM DEQ2015-331722; A.C. by the Northern Portugal Regional Operational Programme (NORTE 2020), under the Portugal 2020 Partnership Agreement, through the European Regional Development Fund (FEDER) (NORTE-01-0145-FEDER-000013), and by the Fundação para a Ciência e Tecnologia (FCT) (IF/00735/2014 and SFRH/BPD/96176/2013).

Author Contributions G.D.B., J.-P.L. and J.A.W. conceived and designed the study and guided the interpretation of the results. A.E.C., M.H.T.S. and V.A. performed the majority of the experiments and data analysis. S.B., D.M.R., P.A., S.E.H., I.M.D., B.K., A.P., J.A.W., C.C., M.d.G.T.S., C.A.W., R.Y. and B.H. conducted experiments and data analysis. Y.L. and T.F. performed the glycan microarray experiments and data analysis. M.J., M.G.N., F.L.v.d.V., J.F.L., A.Car. and A.Car. provided the human patient data and analysis. A.A.B., K.J.K.-C., I.V., M.P., C.Z., M.Z. and N.A.R.G. provided critical conceptual input and reagents. G.D.B. drafted the manuscript. All authors discussed the results, edited and approved the draft and final versions of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to G.D.B. (gordon.brown@abdn.ac.uk).

Reviewer Information Nature thanks A. Casadevall, C. Reis e Sousa and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Mice and fungal strains. C57BL/6 and *Clec1a*^{-/-} mice (8–12 weeks old) were obtained from the specific pathogen-free facility at the University of Aberdeen. Animal experiments were performed using age-matched female mice and conformed to the animal care and welfare protocols approved by the UK Home Office (project license 70/8073) in compliance with all relevant local ethical regulations. *Clec1a*^{-/-} mice were generated commercially (TaconicArtemis) by conventional gene targeting in C57BL/6 embryonic stem cells, as detailed in Extended Data Fig. 7. Sample sizes of at least five animals per group were chosen as this would allow the detection of a 25% difference in the mean between experimental and control groups with a probability of greater than 95% ($P < 0.05$), assuming a standard deviation of around 15% and a minimum power value of 0.8. Mice were randomly assigned to experimental or control groups, co-housed, and experiments were not blinded.

A. fumigatus isolate 13073 (American Type Culture Collection (ATCC)) and a clinical isolate CBS 144-89 (ref. 25) were used as wild-type strains. Melanin mutant strains ($\Delta alb1$ (*pksP1*), $\Delta ayg1$, $\Delta abr1$, $\Delta abr2$, $\Delta arp1$ and $\Delta arp2$ in the B5233 wild-type background) were generated previously²⁶. $\Delta rodA$ and $\Delta pksP\Delta rodA$ deletion mutants were generated as described^{9,10,25}. Wild-type CBS110.46 (ref. 27) and CBS386.75 (ref. 27) strains with white conidia were also used, as indicated. All strains were maintained on 2% (w/v) malt-agar slants or potato dextrose agar in culture flasks; conidia were collected and washed before use. Swollen and/or germinating morphotypes were obtained upon incubating conidia in Sabouraud liquid culture medium at 37°C for different time intervals, as indicated.

Fc-MelLec production and immunolabelling. Soluble chimaeric proteins containing the stalk and C-type lectin-like domain of mouse and human MelLec fused to the mutated Fc portion of human immunoglobulin G1 (IgG1) were generated essentially as described previously²⁸. In brief, the relevant portions of the MelLec encoding genes were amplified by PCR (mouse primers, GAATTCCTCGAGCTGGAGCTCTCCAGGTAC and AAGCTTTCTA CCAGCTGTCTAAT; human primers, GGGATCCACTACTACCAGCTCTCC and TGGATATCTGTCACTTCGCCTAATGTTTC), and cloned into the pSecTag2 expression vector (Invitrogen Life Technologies) containing the mutated form of human IgG1 (ref. 29). Sequenced constructs were transfected into HEK293T cells using Fugene 6 (Promega), as per the manufacturer's instructions. Fusion proteins were purified from conditioned supernatants by chromatography on protein A Sepharose and dialysed against PBS. Fc-Dectin-1 (ref. 28) and Fc-CLEC12B (ref. 24) were generated similarly and used as controls.

For flow cytometry, fungi were incubated in 1–3% (w/v) bovine serum albumin (BSA) in PBS and Fc-proteins were added to a final concentration of 5 $\mu\text{g ml}^{-1}$. Following incubation at 4°C, fungal particles were washed in fluorescence-activated cell sorting (FACS) buffer (0.5% (w/v) BSA and 2 mM EDTA in PBS), and bound Fc-proteins detected with allophycocyanin (APC) or phycoerythrin (PE)-conjugated donkey anti-human antibody (Jackson ImmunoResearch), fixed in 1% (v/v) formaldehyde, and analysed.

For microscopy, paraformaldehyde-fixed conidia were taken for immunolabelling⁶. To obtain NaOH-treated conidia, 1 M NaOH was added to conidia and heated in a boiling water bath for 1 h, followed by centrifugation to collect conidia and extensive washing. Melanin ghosts were isolated from conidia (B5233 as well as melanin mutant strains) by harsh chemical treatments that degrade other cellular components and result in hollow spherical-shaped melanin shells ('melanin ghosts'), as described previously³⁰. For immunostaining, fungal particles were incubated with 5 $\mu\text{g ml}^{-1}$ Fc-MelLec in PBS containing 1% (w/v) BSA for 1 h. After washing, bound Fc-MelLec was detected with fluorescein isothiocyanate (FITC)-labelled goat anti-human Fc-specific IgG (Sigma) in PBS containing 1% (w/v) BSA and washed with PBS-Tween. Conidia incubated with FITC-labelled anti-human Fc-specific IgG only were used as a negative control (data not shown). Labelled conidia were observed by fluorescence microscopy (Leica DMLB).

In some experiments, Fc-proteins were pretreated with either ghosts of melanin mutant strains, as indicated, or YWA1, which was isolated from *AYG1* deletion mutant conidia as described previously¹³.

Monoclonal antibody production. The generation of monoclonal antibodies to MelLec was performed essentially as described previously³¹. In brief, Sprague Dawley rats were immunized with Fc-MelLec in Freund's complete adjuvant. After a final intraperitoneal boost, without adjuvant, rat splenocytes were collected and fused with Y3 myeloma cells, as described³². Hybridoma supernatants were screened by ELISA and positives were then tested by immunohistochemistry and flow cytometry, as described below, against Fc-MelLec as well as MelLec-transduced NIH3T3 fibroblasts. Two monoclonal antibodies (14C8 and 18E4; both IgG1) were selected for further use. Where required, they were biotinylated with Sulfo-NHS-LC Biotin (Pierce), as described by the manufacturer.

Cell culture and growth conditions. Cells were maintained at 37°C and 5% CO₂ in DMEM or RPMI medium supplemented with 10% heat-inactivated fetal calf serum, 100 units per ml penicillin, 0.1 mg ml⁻¹ streptomycin, and 2 mM L-glutamine. Phoenix ecotropic packaging cells (Plat-E) were maintained with the addition of 1 $\mu\text{g ml}^{-1}$ puromycin and 10 $\mu\text{g ml}^{-1}$ blasticidin (Life Technologies, Inc.).

NIH3T3 fibroblasts stably expressing full-length murine or human MelLec were generated essentially as described for Dectin-1 (ref. 33). In brief, haemagglutinin-tagged MelLec was generated through PCR amplification (mouse primers, GGA TCCACCATGCAGGCCAAATACAGCA and CTCGAGCTACTGGAGCT CTCCAGGTAC; human primers, AAAGGATCCACCATGCAGGCCAAGT ACAGCAGCAC and AGCGTAATCCGGAACATCGTATGGGTACTCGAG) and subcloning into pFb-neo (Stratagene). Constructs were transfected into Plat-E retroviral packaging cell lines using Fugene 6 and retrovirus-containing supernatants were used to transduce NIH3T3 fibroblasts in the presence of polybrene (Sigma). Stably transfected cells were selected using 600 $\mu\text{g ml}^{-1}$ G418 (ThermoFisher Inc.). The expression of the receptor was confirmed by flow cytometry and western blotting, using the anti-haemagglutinin antibody (Covance). In some experiments, NIH3T3 cells expressing murine CLEC12A (ref. 31) were used as controls.

ELISA assays. Alkali-insoluble and soluble fungal cell wall fractions were obtained as described previously³⁴. For the ELISA, wells were coated overnight with 20–200 $\mu\text{g ml}^{-1}$ of the cell wall fractions, melanin ghosts (in-house generated), purified YWA1 (in-house generated), 1,8-DHN, 1,2-DHN, 1,4-DHN, naphthalene, 1-naphthol (all from Sigma) in 50 mM carbonate buffer pH 9.6, and then blocked with 1% BSA in PBS. Fc-MelLec (5 $\mu\text{g ml}^{-1}$) was added to the wells and incubated for 1 h at room temperature. After washing with PBS-Tween-20 (0.5% (v/v)), peroxidase-conjugated human Fc-specific IgG (Sigma) was added to the wells and incubated for 1 h at room temperature. After further washing, quantification of Fc-MelLec binding was detected using *o*-phenylenediamine (Sigma) and H₂O₂ detection system (Merck). The reaction was stopped using 4% (v/v) H₂SO₄ and optical densities were measured at 492 nm.

Expression analysis. Peripheral blood leucocytes, resident and thioglycolate-elicited inflammatory peritoneal cells, alveolar macrophages and bone marrow cells were isolated essentially as described previously³⁵. For platelets, peripheral blood was collected in 3.8% (w/v) sodium citrate buffer and centrifuged at 200g at 25°C. The supernatant, containing platelet-rich plasma, was used for subsequent analysis. Bone marrow-derived macrophages or dendritic cells, generated using L929 conditioned medium or 20 ng ml⁻¹ granulocyte-macrophage colony-stimulating factor (R&D Systems), respectively, were prepared as described³¹. In some experiments, cells were stimulated with 100 ng ml⁻¹ lipopolysaccharide from *Escherichia coli* (Sigma).

Tissues isolated from mice were cut into small pieces and incubated for 30 min at 37°C with Liberase (Roche) and DNase (Roche) in RPMI (Gibco), except for the small intestine which was incubated with Collagenase VIII (Sigma-Aldrich). Cells were disaggregated using the gentleMACS Dissociator (Miltenyi), strained through 70- μm nylon cell strainers (Fisher Scientific) and collected by centrifugation. Red blood cells were removed using Pharm Lyse (BD Biosciences).

For flow cytometry, isolated cells were washed in FACS wash (PBS with 0.5% (w/v) BSA and 5–10 mM EDTA) containing anti-CD16/CD32 (Clone 2.4G2, prepared in house). The following antibodies (all from BD Biosciences, eBioscience or Abcam) were used for FACS analysis of cell surface antigen expression following standard methodology: anti-MelLec-Biotin (described above), Streptavidin-PE-CF594 or Streptavidin-APC, anti-CD45.2-FITC (clone 104), biotinylated anti-CD61 (clone 2C9.G2), anti-CD326-APC (EPCAM; clone G8.8), anti-CD31-PE-Cyanine7 (PECAM-1; clone 390), anti-Ly6G-APC (clone 1A8), anti-CD11b-PE-Cy7 (clone M1/70), CD11c-PerCP-Cy5.5 (clone HL3), anti-Siglec-F-BV421 (clone E50-2440), anti-CD45-PE-Cy7 (clone 30-F11), anti-CD11b-PerCP-Cy5.5 (clone M1/70), anti-CD11c-BV421 (clone HL3), anti-Siglec-F-PE (clone E50-2440) and anti-F4/80-AF700 (clone CL:A3-1) and isotype control AFRC MAC 49 (ECACC 85060404; isotype for anti-MelLec). Cell viability was detected using the fixable viability dye eFluor-780 (eBioscience) in PBS, and the cells were fixed with 1% formaldehyde (v/v) before acquisition on a LSRII, LSR Fortessa or FACS Calibur (Becton Dickinson). Data were analysed using FlowJo. All contour plots were constructed using 2% probability contouring with outliers.

For immunofluorescence microscopy, 6 μm alcohol-fixed frozen lung sections were treated with Liberase (Roche) for 4 min at room temperature. Sections were blocked with 2% (v/v) normal goat serum, and incubated with anti-MelLec or isotype control for 1 h followed by AlexaFluor 488 goat anti-rat (Invitrogen) for 30 min. Vectashield with DAPI or propidium iodide (Vector Laboratories Inc.) was used as a mountant for fluorescence, and slides were visualized using a Zeiss LSM 700 confocal microscope. Visualization of transfected NIH3T3 cells was performed similarly, except the cells were fixed in 4% (v/v) paraformaldehyde before analysis.

Detection of MelLec in complementary DNA (cDNA) (Multiple Tissue Panels, Clontech) was performed using the Titanium Taq PCR kit (Clontech) with the following primers: CAGAGCCAGGCACTCAGAGAATG and TGCGGGAGAGCCCTGTCCAAT. Expression of G3PDH was detected as described previously³⁶.

Mouse infection models. For pulmonary infections, 10^5 (corticosteroid model) or 10^7 *A. fumigatus* ATCC 13073 conidia were administered to the caudal oropharynx of anaesthetized mice. In some experiments, mice were administered with 0.6 mg of the corticosteroid, triamcinolone acetonide³⁷ (Bristol-Myers Squibb), on days -1, 1 and 3, relative to infection on day 0. For systemic infections, 10^6 *A. fumigatus* ATCC 13073 conidia were administered to the lateral tail vein of mice. Mice were killed when they had lost 30% body weight or had become moribund. Pulmonary cellular inflammation was assessed by flow cytometry after total lung enzymatic digest, as described above, or in bronchoalveolar lavage samples, isolated with PBS containing 5 mM EDTA (Gibco). Organs were homogenized in PBS and used for the determination of fungal burdens and levels of inflammatory cytokines. Fungal burdens were determined by serial dilution onto potato dextrose agar plates and normalized to organ weights. Cytokines were measured by ELISA (R&D DuoSet), as described by the manufacturer, and normalized to protein concentration. Bone marrow chimaeric mice were generated as described previously³⁸.

Glycan microarray analyses. Microarray analyses were carried out using the neoglycolipid-based microarray system³⁹. Details of the glycan probe library, the generation of the microarrays, imaging and data analysis are in the supplementary glycan microarray document (Supplementary Table 3) in accordance with the MIRAGE (minimum information required for a glycomics experiment) guidelines for reporting glycan microarray-based data. The microarray contained 496 lipid-linked glycan probes (Supplementary Table 1). Microarray analysis of the soluble Fc-MelLec was performed essentially as described⁸. In brief, after blocking arrayed slides with 0.04% (v/v) Blocker Casein (Pierce), 1% (w/v) BSA (Sigma A8577) in HEPES-buffered saline (5 mM HEPES, pH 7.4, 150 mM NaCl, 5 mM CaCl_2), mouse Fc-MelLec was precomplexed with biotinylated anti-human IgG (Vector) at a 1:3 ratio (w/w) before application onto the slides at a final concentration of $10 \mu\text{g ml}^{-1}$. Flag-tagged human langerin was included as a positive control. The protein and rat anti-Flag antibody mAb L5 were provided by C. G. Park. For analysis, the Flag-tagged langerin was precomplexed with the rat anti-Flag at a 1:3 ratio (w/w) and applied onto the slides at a final concentration of $5 \mu\text{g ml}^{-1}$, followed by biotinylated anti-rat IgG (Pierce) ($5 \mu\text{g ml}^{-1}$). To detect binding, AlexaFluor-647-labelled streptavidin from Molecular Probes was used at $1 \mu\text{g ml}^{-1}$. Data analysis and presentation was performed with dedicated glycan microarray software⁴⁰.

Human studies. A total of 310 haematologic patients undergoing allogeneic haematopoietic stem cell transplantation at the Hospital of Santa Maria, Lisbon and Instituto Português de Oncologia (IPO), Porto, between 2009 and 2014 were enrolled in the study. The cases of invasive aspergillosis were identified and classified as 'probable' or 'proven' according to the revised standard criteria from the European Organization for Research and Treatment of Cancer/Mycology Study Group (EORTC/MSG)⁴¹. Exclusion criteria included diagnosis of 'possible' invasive aspergillosis, infection with invasive moulds other than *Aspergillus* spp. or history of pre-transplant mould infection. Study approval was obtained from the institutional review boards (SECVS-125/2014, HSM-632/14 and CES.26/015) and from the National Data Protection Commission (CNPd, 1950/2015) and was in compliance with all local relevant ethical regulations.

Genomic DNA was isolated from whole blood of recipients and donors (before transplantation) using the QIAcube automated system (Qiagen) at the regional centres of the Instituto Português do Sangue e Transplantação (Portugal). Genotyping of the nonsynonymous rs2306894 SNP in the *CLEC1A* gene was performed using KASPar assays (LGC Genomics) according to the manufacturer's instructions in an Applied Biosystems 7500 Fast real-time PCR system (Thermo Fisher). Genotyping sets included randomly selected replicates of previously typed samples, and agreement between original and duplicate samples was $\geq 99\%$.

Peripheral blood mononuclear cells from healthy genotyped donors were enriched from buffy coats using Histopaque-1077 (Sigma-Aldrich) and contaminating erythrocytes were removed using Red Blood Cell Lysis Buffer (Sigma-Aldrich). Participants gave written informed consent before blood collection. Monocytes were isolated by positive selection using magnetically labelled CD14⁺ MicroBeads (Miltenyi Biotec) on a MiniMACS separator and seeded at 10^6 cells per ml in 24-well plates for 7 days in RPMI-1640 medium supplemented with 10% (v/v) human serum and 20 ng ml^{-1} recombinant human granulocyte macrophage colony-stimulating factor (GM-CSF, Gibco). Acquisition of macrophage morphology was confirmed by phase contrast microscopy (Axiovert 135, Zeiss). For infection, macrophages were washed and then infected with live conidia of *A. fumigatus* strain A1163 at a ratio of 1:10 (cells:fungus) for 20 h at 37°C and 5% CO_2 . Cytokines in supernatants were detected using DuoSet ELISA systems

(BioLegend), according to the manufacturer's instructions. At least two technical replicates were performed for each donor.

For the independent cohort, genomic DNA was isolated from EDTA venous blood of healthy Dutch volunteers using the Gentra Pure Gene Blood kit (Qiagen) and genotyped for *CLEC1A* polymorphisms using the Illumina ImmunoChip SNP array platform, described previously⁴². Participants gave written informed consent before blood collection. As the *CLEC1A* SNP of interest (exonic rs2306894) was not represented on the genotyping platform, two intronic polymorphisms (rs7972187 and rs3825300) were used as markers. Linkage analysis revealed complete linkage disequilibrium of these two polymorphisms with the SNP of interest⁴³ ($R^2 = 1$). A total of 10^5 peripheral blood mononuclear cells isolated from genotyped donors were stimulated with 10^7 killed conidia per ml of the clinical isolate *A. fumigatus* V05-27 (ref. 44). After 24-h incubation in the presence of 10% human pooled serum at 37°C and 5% CO_2 , supernatants were collected and IL-1 β and IL-8 were measured by ELISA (R&D Systems and PeliKine, respectively).

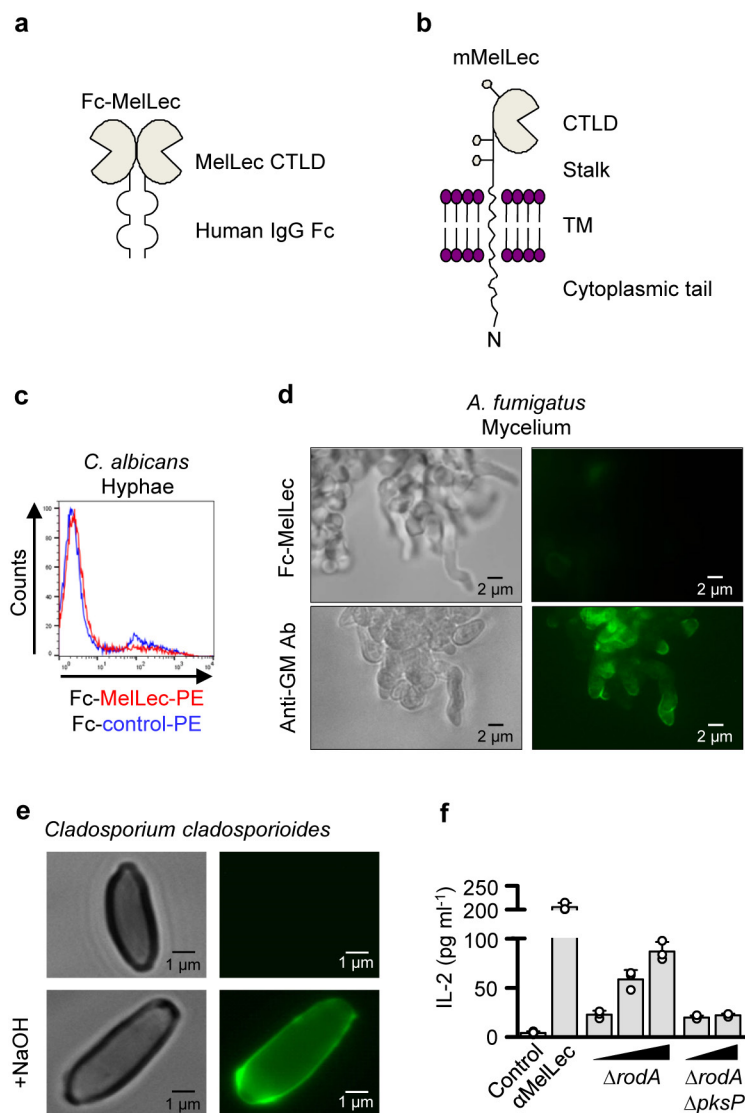
Statistical analysis. The probability of invasive aspergillosis resulting from *CLEC1A* rs2306894 SNP was analysed using the cumulative incidence method and compared using Gray's test⁴⁵. Cumulative incidences were computed with the cmprsk package for R version 2.10.1 (ref. 46), with censoring of data at the date of last follow-up visit and defining relapse and death as competing events. A period of 24 months after transplant was chosen to include all cases of fungal infection. Mouse survival data were analysed with the log rank test using GraphPad Prism. No data was excluded.

In vitro and *ex vivo* data were analysed using the GraphPad Prism software. Two-tailed student's *t*-tests or Mann-Whitney *U* tests were used to determine statistical significance. All experiments were independently repeated at least once, unless otherwise indicated.

Data availability. The data that support the findings of this study are available from the corresponding author upon reasonable request.

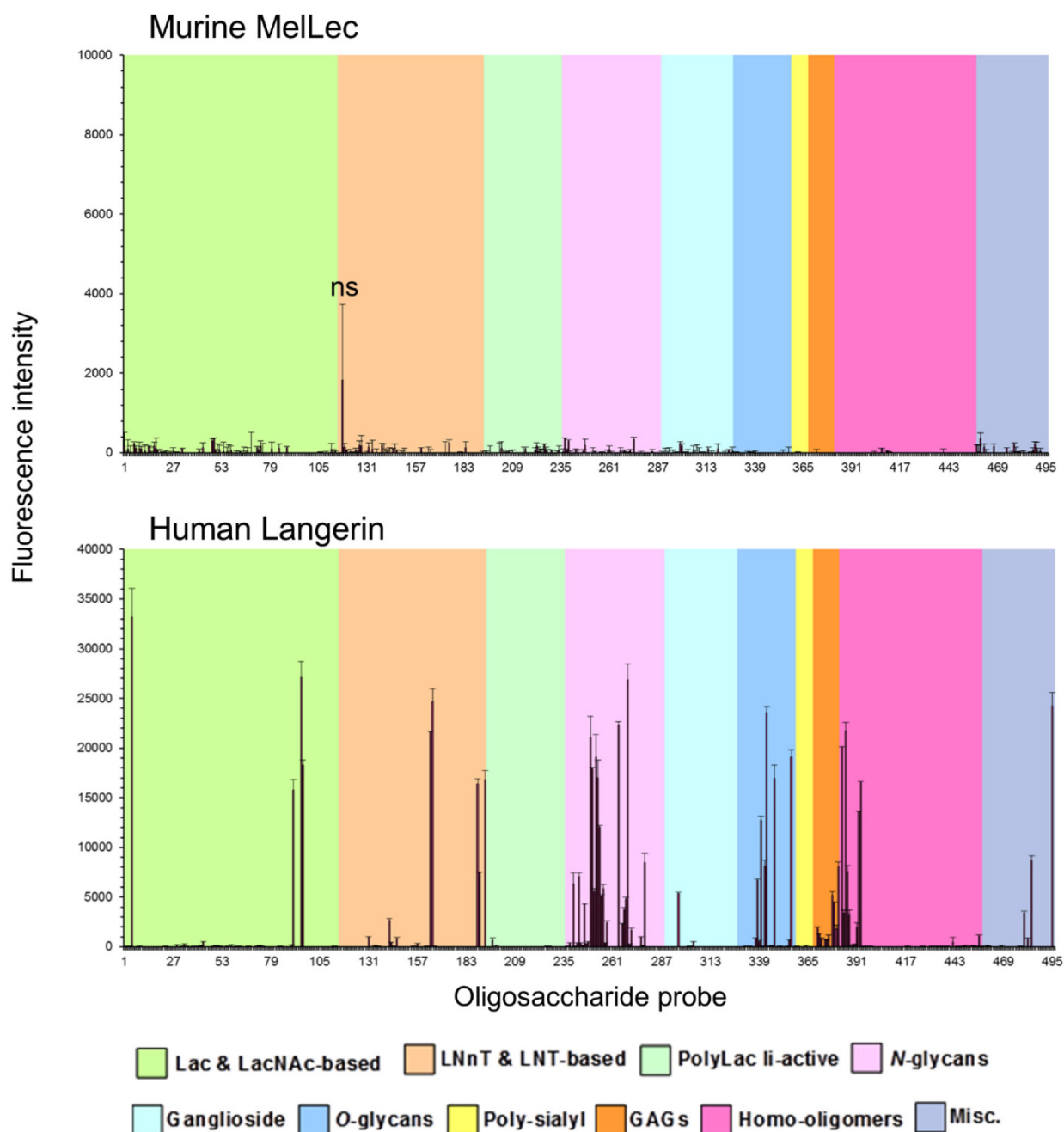
25. Thau, N. *et al.* rodletless mutants of *Aspergillus fumigatus*. *Infect. Immun.* **62**, 4380–4388 (1994).
26. Tsai, H. F., Wheeler, M. H., Chang, Y. C. & Kwon-Chung, K. J. A developmentally regulated gene cluster involved in conidial pigment biosynthesis in *Aspergillus fumigatus*. *J. Bacteriol.* **181**, 6469–6477 (1999).
27. Sarfati, J. *et al.* A new experimental murine aspergillosis model to identify strains of *Aspergillus fumigatus* with reduced virulence. *Nippon Ishinkin Gakkai Zasshi* **43**, 203–213 (2002).
28. Graham, L. M. *et al.* Soluble Dectin-1 as a tool to detect β -glucans. *J. Immunol. Methods* **314**, 164–169 (2006).
29. Ettinger, R., Browning, J. L., Michie, S. A., van Ewijk, W. & McDevitt, H. O. Disrupted splenic architecture, but normal lymph node development in mice expressing a soluble lymphotoxin- β receptor-IgG1 fusion protein. *Proc. Natl Acad. Sci. USA* **93**, 13102–13107 (1996).
30. Bayry, J. *et al.* Surface structure characterization of *Aspergillus fumigatus* conidia mutated in the melanin synthesis pathway and their human cellular immune response. *Infect. Immun.* **82**, 3141–3153 (2014).
31. Pyz, E. *et al.* Characterisation of murine MICL (CLEC12A) and evidence for an endogenous ligand. *Eur. J. Immunol.* **38**, 1157–1163 (2008).
32. Galfré, G., Milstein, C. & Wright, B. Rat \times rat hybrid myelomas and a monoclonal anti-Fd portion of mouse IgG. *Nature* **277**, 131–133 (1979).
33. Willment, J. A., Gordon, S. & Brown, G. D. Characterization of the human β -glucan receptor and its alternatively spliced isoforms. *J. Biol. Chem.* **276**, 43818–43823 (2001).
34. Richie, D. L. *et al.* A role for the unfolded protein response (UPR) in virulence and antifungal susceptibility in *Aspergillus fumigatus*. *PLoS Pathog.* **5**, e1000258 (2009).
35. Taylor, P. R., Brown, G. D., Geldhof, A. B., Martinez-Pomares, L. & Gordon, S. Pattern recognition receptors and differentiation antigens define murine myeloid cell heterogeneity *ex vivo*. *Eur. J. Immunol.* **33**, 2090–2097 (2003).
36. Kerscher, B. *et al.* Mycobacterial receptor, Clec4d (CLECSF8, MCL), is coregulated with Mincle and upregulated on mouse myeloid cells following microbial challenge. *Eur. J. Immunol.* **46**, 381–389 (2016).
37. Shepardson, K. M. *et al.* Myeloid derived hypoxia inducible factor 1-alpha is required for protection against pulmonary *Aspergillus fumigatus* infection. *PLoS Pathog.* **10**, e1004378 (2014).
38. Redelinghuys, P. *et al.* MICL controls inflammation in rheumatoid arthritis. *Ann. Rheum. Dis.* **75**, 1386–1391 (2016).
39. Liu, Y. *et al.* Neoglycolipid-based oligosaccharide microarray system: preparation of NGLs and their noncovalent immobilization on nitrocellulose-coated glass slides for microarray analyses. *Methods Mol. Biol.* **808**, 117–136 (2012).
40. Stoll, M. S. & Feizi, T. Software tools for storing, processing and displaying carbohydrate microarray data. In *Proc. Beilstein Symposium on Glyco-Bioinformatics* (ed. Kettner, C.) 123–140 (Beilstein, 2009).
41. De Pauw, B. *et al.* Revised definitions of invasive fungal disease from the European Organization for Research and Treatment of Cancer/Invasive Fungal Infections Cooperative Group and the National Institute of Allergy and Infectious Diseases Mycoses Study Group (EORTC/MSG) Consensus Group. *Clin. Infect. Dis.* **46**, 1813–1821 (2008).

42. Kumar, V. *et al.* Immunochip SNP array identifies novel genetic variants conferring susceptibility to candidaemia. *Nat. Commun.* **5**, 4675 (2014).
43. Johnson, A. D. *et al.* SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* **24**, 2938–2939 (2008).
44. Netea, M. G. *et al.* *Aspergillus fumigatus* evades immune recognition during germination through loss of toll-like receptor-4-mediated signal transduction. *J. Infect. Dis.* **188**, 320–326 (2003).
45. Gray, R. T. A class of K -sample tests for comparing the cumulative incidence of a competing risk. *Ann. Stat.* **16**, 1141–1154 (1988).
46. Scrucca, L., Santucci, A. & Aversa, F. Competing risk analysis using R: an easy guide for clinicians. *Bone Marrow Transplant.* **40**, 381–387 (2007).
47. LeibundGut-Landmann, S., Osorio, F., Brown, G. D. & Reis e Sousa, C. Stimulation of dendritic cells via the dectin-1/Syk pathway allows priming of cytotoxic T-cell responses. *Blood* **112**, 4971–4980 (2008).



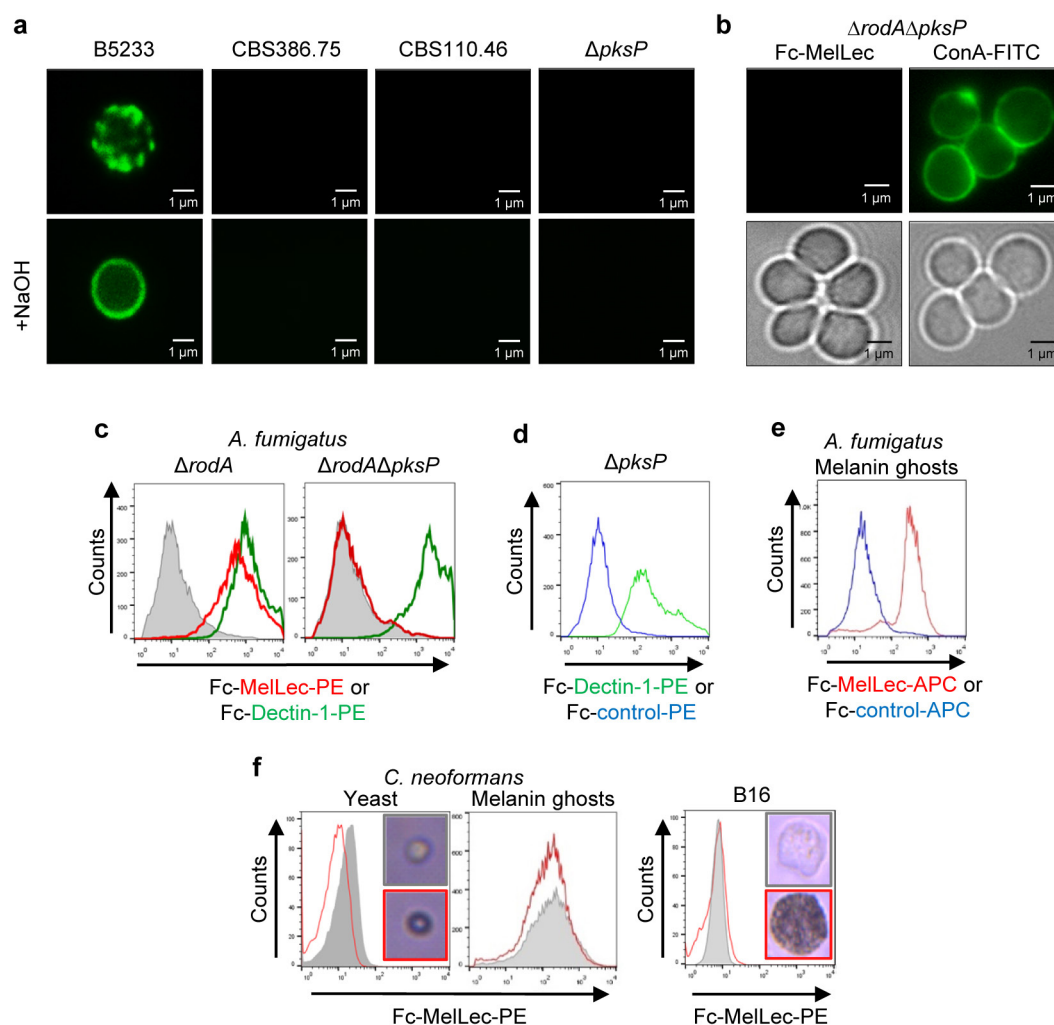
Extended Data Figure 1 | MelLec recognizes ligands on selected fungi and fungal morphotypes. a, b, Cartoon representations of the structures of Fc-MelLec (**a**) and the full-length receptor (**b**). Lollipop structures represent predicted glycosylation sites. **c,** Fc-MelLec or Fc-CLEC12B (Fc-control) staining of *C. albicans* hyphae, generated in RPMI with 10% fetal bovine serum for 90 min. Fungal particles were analysed by flow cytometry. Experiment was repeated independently twice, with similar results. **d,** Representative light microscopy images and immunofluorescence micrographs using Fc-MelLec or anti-galactomannan (GM, control) as probes to detect ligands on *A. fumigatus* mycelium. Experiments were repeated three times independently,

with similar results. **e,** Representative light microscopy images and immunofluorescence micrographs showing the surface distribution of MelLec ligands on *C. cladosporioides* using Fc-MelLec as a probe. The lower panels show fungal cells after treatment with 1 M NaOH. Experiments were repeated three times independently, with similar results. **f,** IL-2 production by MelLec-expressing BWZ reporter cells after stimulation by anti-MelLec antibody (α MelLec) crosslinking or with $\Delta rodA$ (1:1, 5:1, 10:1) or $\Delta rodA \Delta pksP$ (5:1, 10:1) *A. fumigatus* conidia, as indicated. Values shown are mean \pm s.d. Experiment was repeated three times independently, with similar results.



Extended Data Figure 2 | Glycan microarray analyses of mouse Fc-MelLec and human langerin. These 496 lipid-linked probes are arranged according to their backbone sequences as annotated in the coloured panels below the figure. GAGs, glycosaminoglycans; Lac, lactose; LacNAc, *N*-acetyllactosamine; LNT, lacto-*N*-tetraose; LNT, lacto-*N*-tetraose;

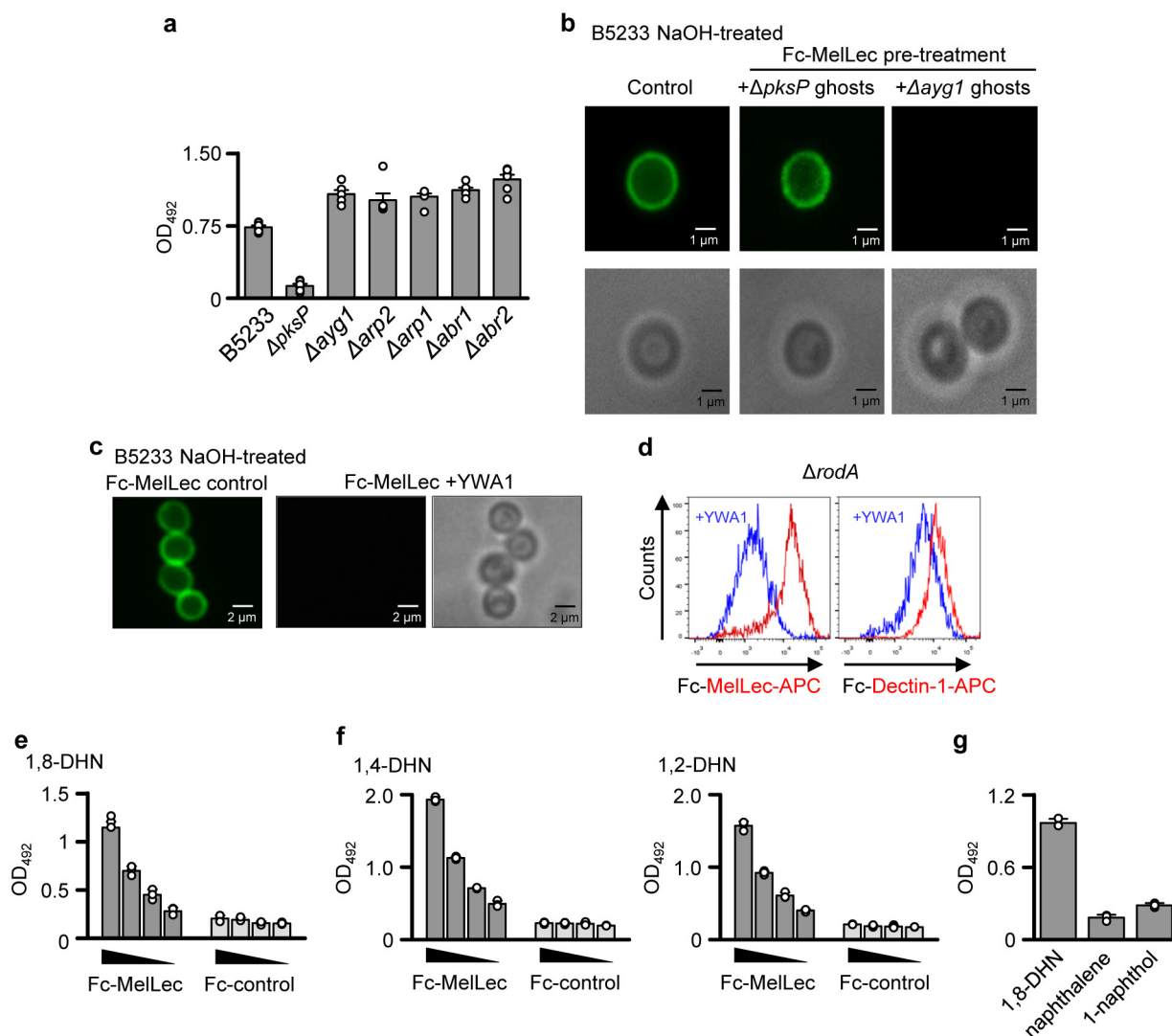
Misc., miscellaneous; PolyLac, polylectosamine. The signals are means of fluorescence intensities of duplicate spots, printed at 5 fM per spot level with error bars representing half of the difference between the two values. The signals shown together with the probe sequences are in Supplementary Table 1. NS, not significant.



Extended Data Figure 3 | MelLec recognizes DHN-melanin.

a, Representative immunofluorescence micrographs using Fc-MelLec as a probe showing the surface distribution of MelLec-ligands on wild-type *A. fumigatus* and various melanin-deficient mutants. Lower panels show conidia treated with 1 M NaOH. **b**, Representative immunofluorescence micrographs and light microscopy images of $\Delta rodA \Delta pksP$ *A. fumigatus* conidia stained with Fc-MelLec or ConA-FITC. **c**, Representative histograms showing the presence or absence of MelLec or Dectin-1 ligands on $\Delta rodA$ or $\Delta rodA \Delta pksP$ *A. fumigatus* conidia. Fungal particles were stained with Fc-MelLec (red) or Fc-Dectin-1 (green) and analysed by flow cytometry. Grey histograms indicate secondary-only control.

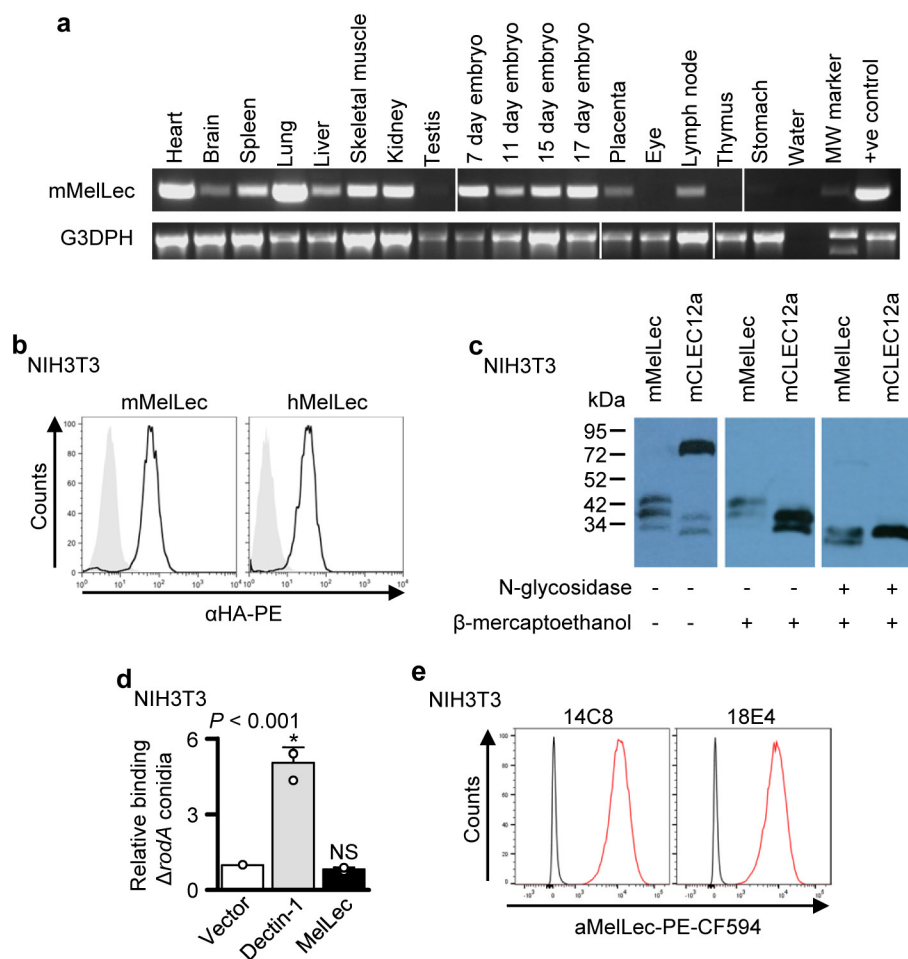
d, Representative histogram showing the presence of Dectin-1 ligands on $\Delta pksP$ *A. fumigatus* conidia. Fungal particles were stained with Fc-Dectin-1 (green) or Fc-CLEC12B (Fc-control; blue) and analysed by flow cytometry. **e**, Representative histogram showing the presence of MelLec ligands on melanin ghosts of *A. fumigatus* conidia. Fungal particles were stained with Fc-MelLec (red) or Fc-CLEC12B (Fc-control; blue) and analysed by flow cytometry. In **a–e**, experiments were repeated three times independently, with similar results. **f**, Flow-cytometric analysis of melanized (red) and non-melanized (grey) *Cryptococcus neoformans* yeast and melanin ghosts, and B16 melanoma cells⁴⁷, stained with Fc-MelLec. The experiment was repeated independently twice, with similar results.



Extended Data Figure 4 | MelLec recognizes naphthalene-diol.

a, Detection of MelLec ligands in ghosts of melanin mutants of *A. fumigatus* by ELISA, as indicated. Values show mean \pm s.d.
b, Representative immunofluorescence and light micrograph images of Fc-MelLec ligands on NaOH-treated *A. fumigatus* B5233 conidia after pretreatment with ghosts of $\Delta pksP$ or $\Delta ayg1$ conidia, as indicated.
c, Representative immunofluorescence micrographs and light microscopy images of MelLec ligands on NaOH-treated wild-type *A. fumigatus* conidia after pretreatment with or without YWA1. In **a–c**, experiments were

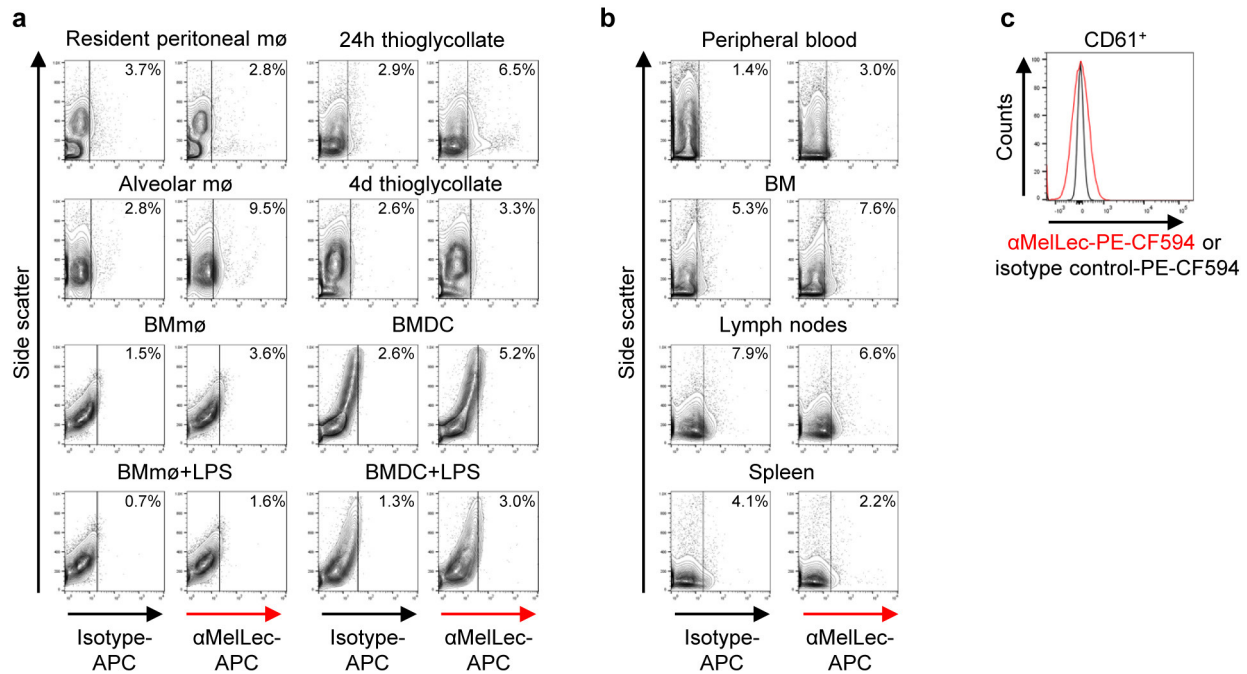
repeated at least three times independently, with similar results.
d, Detection of Fc-MelLec or Fc-Dectin-1 ligands on $\Delta rodA$ conidia after pretreatment with (blue) or without (red) YWA1. The experiment was repeated independently twice, with similar results.
e, f, Detection of 1,8-DHN (**e**) and 1,2-DHN and 1,4-DHN (**f**) by Fc-MelLec and Fc-control using ELISA. Values show mean \pm s.d.
g, Detection of 1,8-DHN, naphthalene and 1-naphthol by Fc-MelLec using ELISA. Values show mean \pm s.d. In **e–g**, experiments were repeated at least three times independently, with similar results.



Extended Data Figure 5 | MelLec is expressed at the cell surface.

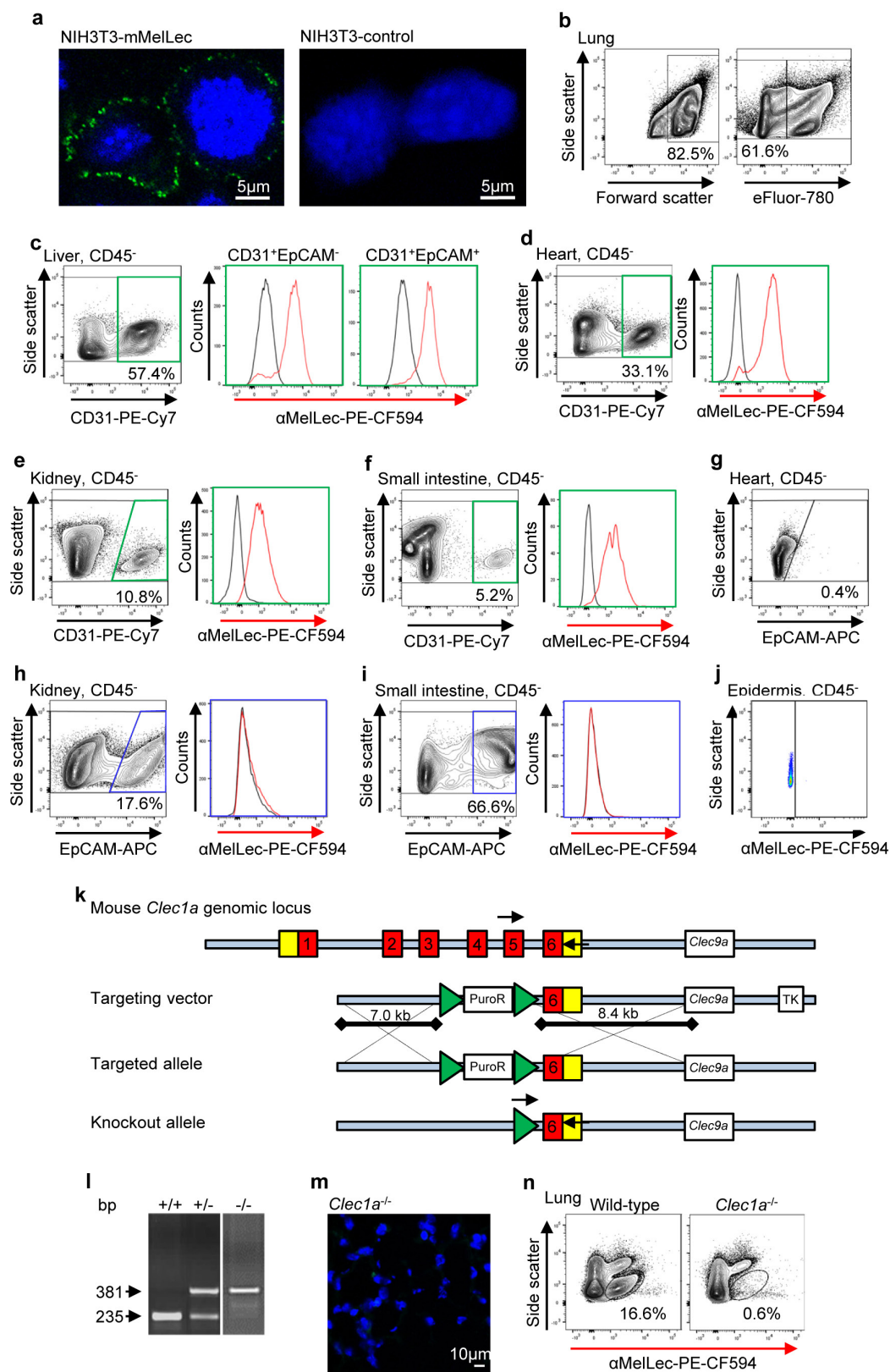
a, RT-PCR detection of MelLec expression in various tissues. The expression of glyceraldehyde-3-phosphate dehydrogenase (G3PDH) in these samples, also used for the characterization of MCL³⁶, is shown as a control. The experiment was performed once; for gel source data, see Supplementary Fig. 1. **b**, Flow-cytometric analysis of surface expression of haemagglutinin-tagged mouse (m) and human (h) MelLec on the surface of NIH3T3 fibroblasts (black open histograms). NIH3T3 cells transfected with vector only served as controls (grey filled histograms). **c**, Western-blot analysis of lysates of haemagglutinin-tagged MelLec expressing

NIH3T3 cells under reducing and non-reducing conditions and with (+) and without (–) *N*-glycosidase. Haemagglutinin-tagged CLEC12A (ref. 31) expressing NIH3T3 cells served as controls (for blot source data, see Supplementary Fig. 1). **d**, Relative binding of FITC-labelled $\Delta rodA$ *A. fumigatus* conidia to NIH3T3 cells transduced with vector only, Dectin-1 or MelLec, as determined by flow cytometry. Values shown are mean \pm s.d., analysed by one-way ANOVA. **e**, Screening of hybridoma supernatants on MelLec-expressing (red) and parental (black) NIH3T3 cells. In **b–e**, experiments were repeated at least three times independently, with similar results. * $P \leq 0.05$; NS, not significant.



Extended Data Figure 6 | Mouse MelLec is not expressed by myeloid cells. **a, b**, Flow-cytometric analysis of MelLec expression on various *ex vivo* and *in vitro* derived myeloid cells (**a**) and peripheral blood, bone marrow, lymph nodes and spleen (**b**). Experiments were repeated at least

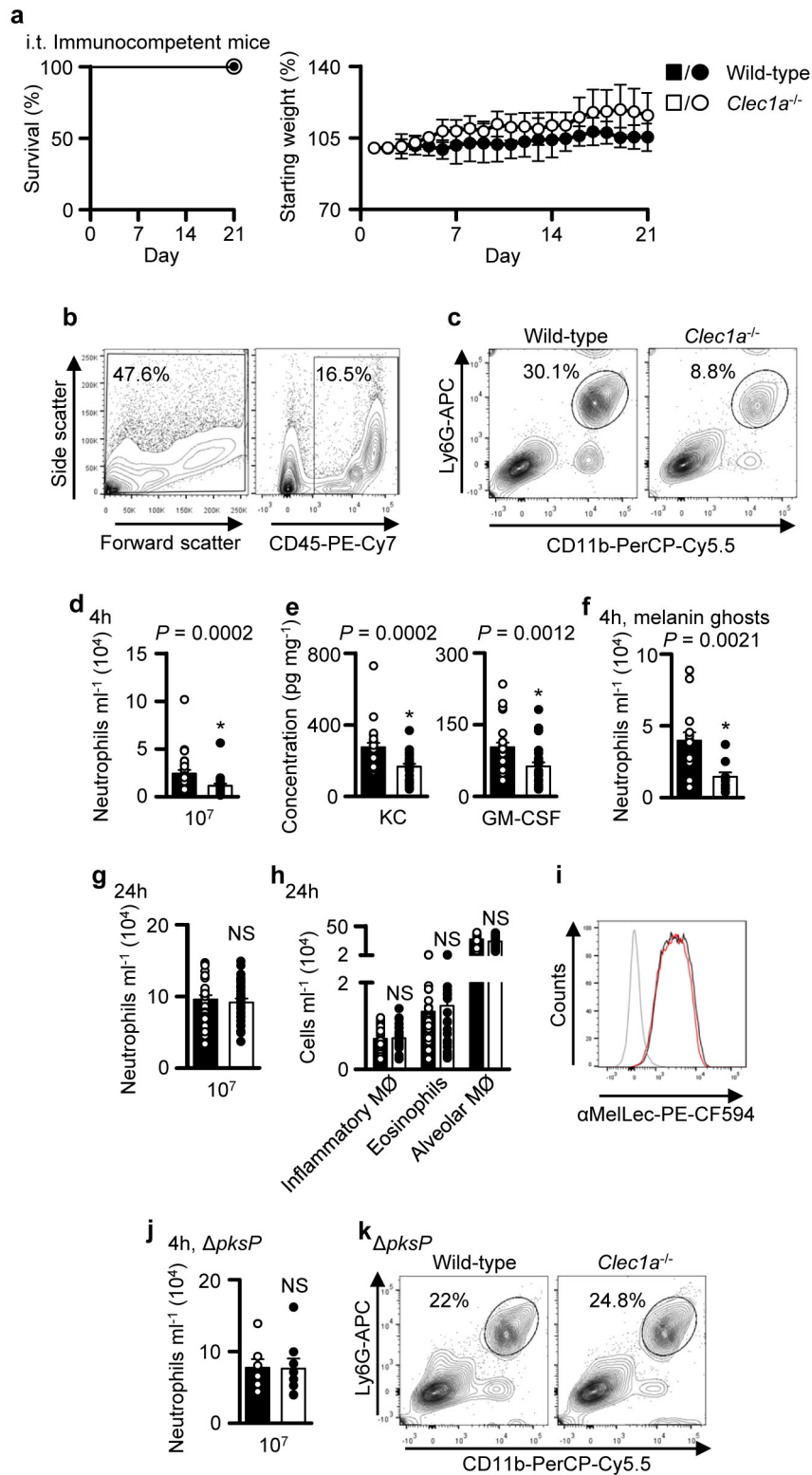
twice independently, with similar results. **c**, Flow-cytometric analysis of MelLec expression on CD61⁺ platelets. Experiment was repeated at least three times independently, with similar results. BM, bone marrow; DC, dendritic cell; LPS, lipopolysaccharide; m ϕ , macrophage.



Extended Data Figure 7 | See next page for caption.

Extended Data Figure 7 | MelLec expression in tissues and generation of *Clec1a*^{-/-} mice. **a**, Immunofluorescence microscopy of MelLec-expressing versus control NIH3T3 cells labelled with anti-MelLec antibody (green). Nuclei are stained with DAPI (blue). Experiments were repeated at least three times independently, with similar results. **b**, Exemplar flow-cytometric gating strategy for identification of live cells from tissue. **c**, Flow-cytometric analysis of MelLec expression on live CD45⁻CD31⁺ EpCAM⁻ and EpCAM⁺ populations in the liver. **d–j**, Flow cytometric analysis of MelLec expression on live CD45⁻CD31⁺ cells in the heart (**d**), kidney (**e**) and small intestine (**f**), and on live CD45⁻EpCAM⁺ cells in the heart (**g**), kidney (**h**), small intestine (**i**) and epidermis (**j**). In **b–j**,

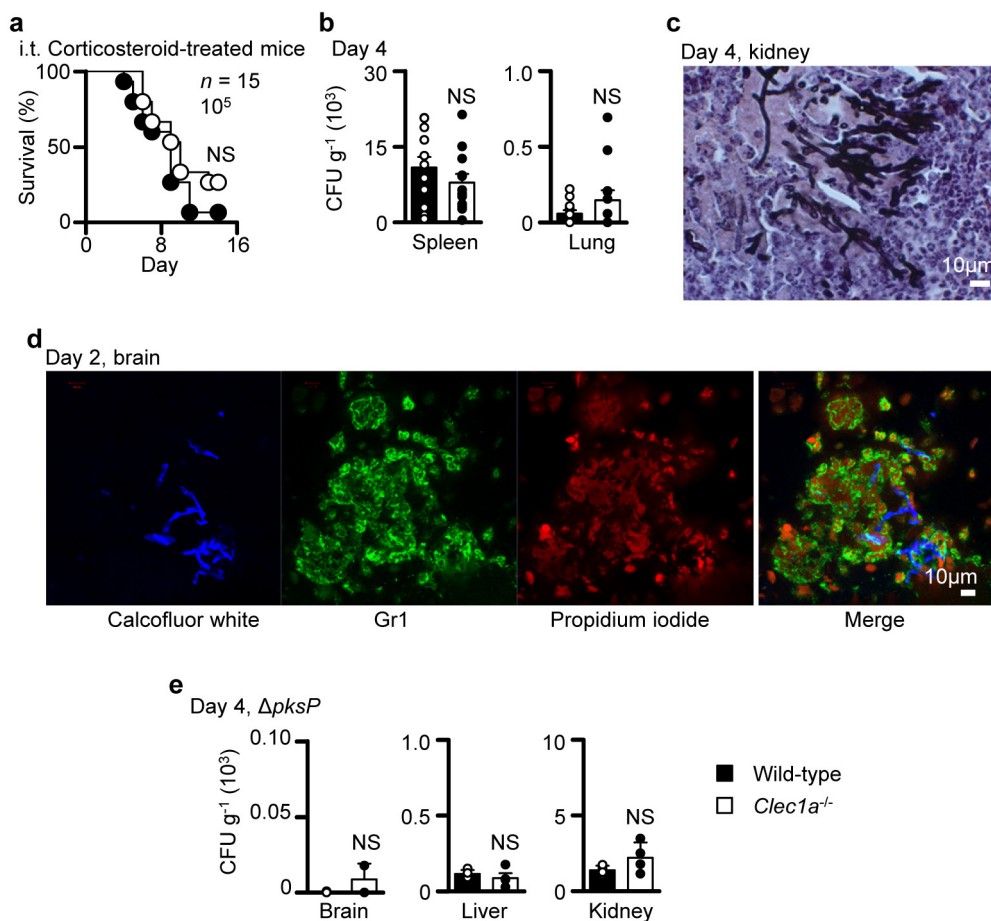
experiments were repeated at least twice independently, with similar results. Black lines, isotype controls. **k**, Schematic of the wild-type *Clec1a* locus, gene targeting vector, PCR primer sites and correctly targeted recombinant allele. **l**, PCR analysis of gene-targeted mice (for gel source data, see Supplementary Fig. 1). +/+ , wild-type, +/- heterozygous and -/- homozygous for the targeted allele. **m**, Immunofluorescence microscopy of naive lung tissue from *Clec1a*^{-/-} mice (labelling of wild-type lung is shown in Fig. 3b). **n**, Analysis of MelLec expression in disaggregated lung tissue from wild-type (wt) or *Clec1a*^{-/-} mice by flow cytometry. In **l–n**, experiments were repeated at least three times independently, with similar results.



Extended Data Figure 8 | See next page for caption.

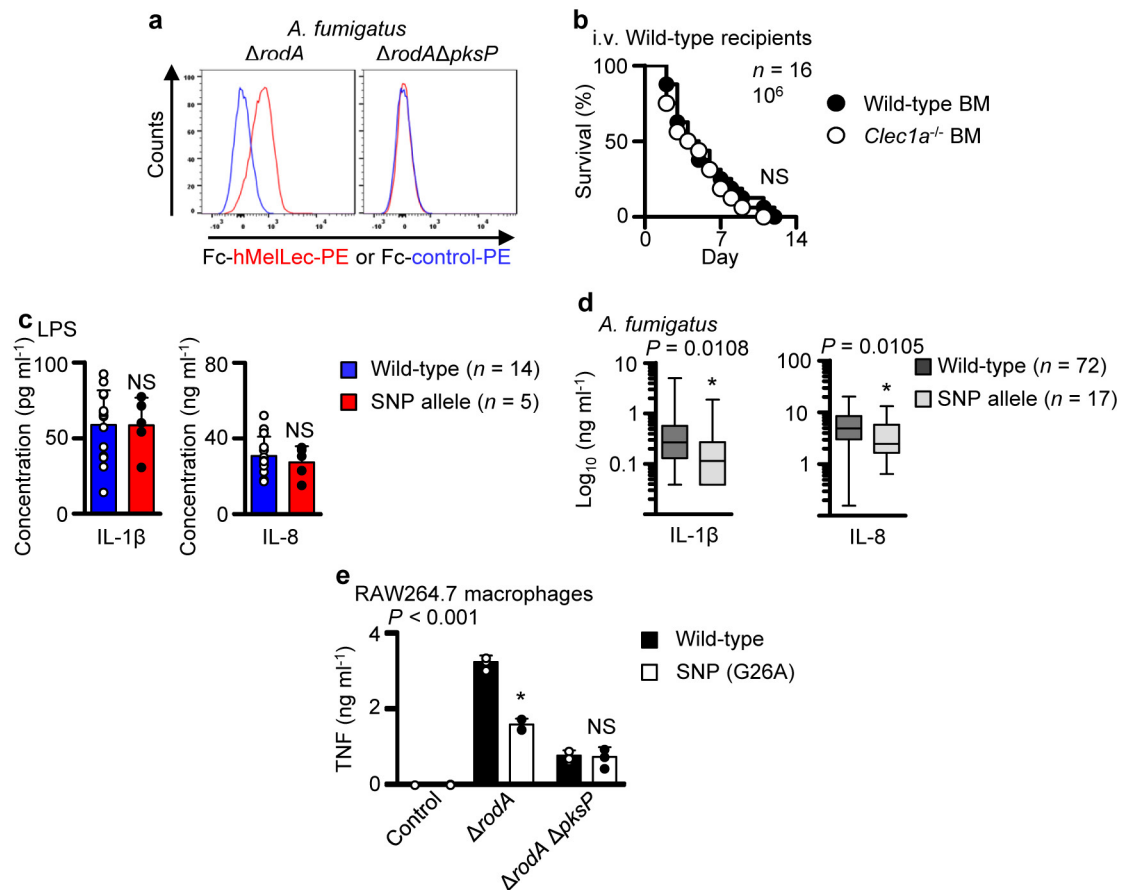
Extended Data Figure 8 | *Clec1a*^{-/-} mice show early inflammatory defects upon challenge with *A. fumigatus*. **a**, Survival (left) and weight measurements (right) of immunocompetent mice after i.t. infection with 10⁷ *A. fumigatus* conidia (*n* = 4 mice per group). Values shown are mean ± s.d. **b**, Exemplar flow-cytometric gating strategy for the identification of CD45⁺ cells from bronchoalveolar lavage. **c**, Representative FACS profiles of pulmonary CD11b⁺ Ly6G^{high} neutrophils in wild-type and *Clec1a*^{-/-} mice 4 h after challenge with 10⁷ *A. fumigatus* conidia (wild-type *n* = 29 mice; *Clec1a*^{-/-} *n* = 26 mice). **d, e**, Pulmonary CD11b⁺ Ly6G^{high} neutrophils (wild-type *n* = 29 mice, *Clec1a*^{-/-} *n* = 26 mice) (**d**) and cytokines (*n* = 25 mice per group) (**e**) in mice 4 h after challenge with 10⁷ *A. fumigatus* conidia. **f**, Pulmonary CD11b⁺ Ly6G^{high} neutrophils in mice 4 h after challenge with melanin ghosts (160 µg) of *A. fumigatus* (wild-type *n* = 15 mice; *Clec1a*^{-/-} *n* = 10 mice). Samples with blood contamination were excluded. **g**, Pulmonary CD11b⁺ Ly6G^{high} neutrophils in mice 24 h after challenge with 10⁷ *A. fumigatus* conidia (wild-type *n* = 33 mice, *Clec1a*^{-/-} *n* = 30 mice).

h, Cellular inflammatory profiles of mice 24 h after challenge with 10⁷ *A. fumigatus* conidia (wild-type *n* = 33 mice, *Clec1a*^{-/-} *n* = 30 mice). Alveolar macrophages were defined as CD11c⁺ Siglec-F⁺, inflammatory macrophages as CD11b⁺ F4/80⁺ and eosinophils as CD11b⁺ Siglec-F⁺. In **d–h**, values shown are mean ± s.e.m. of pooled data from at least two independent experiments, analysed by two-sided Mann–Whitney *U* test. **i**, Expression of MelLec on pulmonary CD45⁺ CD31⁺ cells isolated from uninfected mice (black) and mice 24 h after infection with *A. fumigatus* conidia (red). Grey line, isotype control (*n* = 3 mice per group). **j**, Pulmonary CD11b⁺ Ly6G^{high} neutrophils in mice 4 h after challenge with 10⁷ Δ *pksP* *A. fumigatus* conidia (wild-type *n* = 7 mice, *Clec1a*^{-/-} *n* = 8 mice). Values shown are mean ± s.e.m. of pooled data from two independent experiments, analysed by two-sided Mann–Whitney *U* test. **k**, Representative FACS profiles of pulmonary CD11b⁺ Ly6G^{high} neutrophils in wild-type and *Clec1a*^{-/-} mice 4 h after challenge with 10⁷ Δ *pksP* *A. fumigatus* conidia (wild-type *n* = 7 mice, *Clec1a*^{-/-} *n* = 8 mice). **P* ≤ 0.05; NS, not significant.



Extended Data Figure 9 | $Clec1a^{-/-}$ mice show alterations in antifungal immunity during systemic infection. **a**, Survival of corticosteroid-treated mice after i.t. infection with 10^5 *A. fumigatus* conidia ($n = 15$ mice per group). Pooled data from two independent experiments, analysed by log-rank test. **b**, Fungal burdens in various mouse tissues, as indicated, 4 days after i.v. infection with 10^6 *A. fumigatus* conidia ($n = 12$ mice per group). Values shown are mean \pm s.e.m. of pooled data from two independent experiments, analysed by two-sided Mann–Whitney *U* test. **c**, Tissue section of kidney from day-4-infected $Clec1a^{-/-}$ mouse stained with

Grocott's methenamine silver stain and haematoxylin ($n = 3$ mice per group). **d**, Immunofluorescence microscopy of brain from day-2-infected $Clec1a^{-/-}$ mouse ($n = 3$ mice per group). Fungi are stained with calcofluor white (blue), leukocytes with Gr1 (green) and DNA with propidium iodide (red). **e**, Fungal burdens in various mouse tissues, as indicated, four days after i.v. infection with 10^6 $\Delta pksP$ *A. fumigatus* conidia ($n = 4$ mice per group). Values shown are mean \pm s.d., analysed by two-sided Mann–Whitney *U* test. * $P \leq 0.05$; NS, not significant.



Extended Data Figure 10 | A single nucleotide polymorphism in human MelLec influences anti-*Aspergillus* inflammatory responses.

a, Representative histograms showing the presence or absence of human MelLec ligands on $\Delta rodA$ or $\Delta rodA\Delta pksP$ *A. fumigatus* conidia, as determined by flow cytometry. Fc-CLEC12B was used as a control (Fc-control). The experiment was repeated at least three times independently, with similar results. **b**, Survival of irradiated wild-type mice reconstituted with wild-type or *Clec1a*^{-/-} bone marrow (BM), as indicated, after i.v. infection with 10^6 *A. fumigatus* conidia (n = 16 mice per group). Pooled data from two independent experiments, analysed by log-rank test. **c**, Inflammatory cytokine production in monocyte-derived macrophages isolated from genotyped individuals, after stimulation with lipopolysaccharide (wild-type n = 14 individuals,

SNP allele n = 5 individuals). Values shown are mean \pm s.d., analysed by two-sided Mann–Whitney U test. **d**, Inflammatory cytokine production in peripheral blood mononuclear cells isolated from genotyped Dutch individuals, after stimulation with heat-killed *A. fumigatus* conidia (wild-type n = 72 individuals, SNP allele n = 17 individuals). Boxes represent the median values and interquartile ranges; whiskers represent minimum and maximum values, analysed by two-sided Mann–Whitney U test. **e**, Inflammatory cytokine production in transduced RAW264.7 macrophages expressing wild-type or SNP allele after stimulation with $\Delta rodA$ or $\Delta rodA\Delta pksP$ *A. fumigatus* conidia, as indicated. Values shown are mean \pm s.d., analysed by one-way ANOVA and repeated at least three times independently, with similar results. * $P \leq 0.05$; NS, not significant.

EWS–FLI1 increases transcription to cause R-loops and block BRCA1 repair in Ewing sarcoma

Aparna Gorthi^{1,2}, July Carolina Romero^{1,2}, Eva Loranc², Lin Cao², Liesl A. Lawrence^{1,2}, Elicia Goodale², Amanda Balboni Iniguez^{3,4}, Xavier Bernard^{1,2}, V. Pragathi Masamsetti², Sydney Roston⁵, Elizabeth R. Lawlor⁶, Jeffrey A. Toretzky⁵, Kimberly Stegmaier^{3,4}, Stephen L. Lessnick⁷, Yidong Chen^{2,8,9} & Alexander J. R. Bishop^{1,2,8}

Ewing sarcoma is an aggressive paediatric cancer of the bone and soft tissue. It results from a chromosomal translocation, predominantly t(11;22)(q24;q12), that fuses the N-terminal transactivation domain of the constitutively expressed EWSR1 protein with the C-terminal DNA binding domain of the rarely expressed FLI1 protein¹. Ewing sarcoma is highly sensitive to genotoxic agents such as etoposide, but the underlying molecular basis of this sensitivity is unclear. Here we show that Ewing sarcoma cells display alterations in regulation of damage-induced transcription, accumulation of R-loops and increased replication stress. In addition, homologous recombination is impaired in Ewing sarcoma owing to an enriched interaction between BRCA1 and the elongating transcription machinery. Finally, we uncover a role for EWSR1 in the transcriptional response to damage, suppressing R-loops and promoting homologous recombination. Our findings improve the current understanding of EWSR1 function, elucidate the mechanistic basis of the sensitivity of Ewing sarcoma to chemotherapy (including PARP1 inhibitors) and highlight a class of BRCA-deficient-like tumours.

EWSR1 is an RNA-binding protein that affects RNA metabolism, presumably through its regulation of RNA polymerase II (RNAPII) and coupling with the splicing machinery². There is also evidence that EWSR1 is involved in genome stability³. Despite extensive research on the transcription targets of the fusion of EWSR1 and FLI1 (EWS–FLI1), factors that mediate the chemosensitivity of Ewing sarcoma or the role of EWSR1 have not been well characterized. Compared to control cell lines (Extended Data Fig. 1a), Ewing sarcoma cell lines were acutely sensitive to most forms of damage, including etoposide (topoisomerase II inhibitor) (Fig. 1a). Notably, EWS–FLI1 conferred this chemosensitization (Fig. 1b), beyond the decreased viability caused by knockdown of the oncogene (Extended Data Fig. 1b). Conversely, EWS–FLI1 expression increased chemosensitivity in U2OS osteosarcoma cells (Extended Data Fig. 1c). As independent validation of this finding, the half maximal inhibitory concentration (IC₅₀) of drugs that induced transcription and replication blocks was nearly fivefold lower in EWS–FLI1-associated cancers than in others in a pan-cancer dataset from the Genomics of Drug Sensitivity in Cancer database⁴ (Extended Data Fig. 1d).

Aberrant regulation of transcription is an important source of endogenous DNA damage⁵. To identify pathways that contribute to the chemosensitivity of Ewing sarcoma, we examined gene expression over time after exposure to etoposide. Gene set enrichment analysis contrasting gene expression in Ewing sarcoma and control cells under basal conditions extracted the expected Ewing sarcoma profile along with defects in replication, transcription and repair pathways

(Extended Data Fig. 1e–g). We also identified a subset of genes that were significantly altered in response to damage in IMR90 human lung cells but not Ewing sarcoma (Fig. 1c, Supplementary Table 1); functional annotation analysis revealed significant enrichment for transcription regulation and RNA metabolism genes (Extended Data Table 1a, b). Notably, comparison with genome-wide RNA inhibition (RNAi) survival screens in *Drosophila* Kc167 cells exposed to various damaging agents consistently highlighted RNA metabolism (Extended Data Fig. 1h, Supplementary Table 2), implicating it as a conserved and critical damage survival component.

EWS–FLI1 and EWSR1 are known to interact with each other^{6,7} and with sub-components of the transcriptional machinery^{8,9}. It has previously been suggested that EWS–FLI1 acts in a dominant-negative manner to wild-type EWSR1 in splicing^{7,9}. However, the role of these two proteins in directly controlling RNAPII activity has not been actively studied. The largest subunit of RNAPII is hyperphosphorylated at Ser2 and Ser5 of the heptapeptide repeats in the C-terminal domain (CTD) during active transcription¹⁰; Ser5 phosphorylation (by CDK7/cyclin H) occurs early during initiation and Ser2 phosphorylation (by CDK9/cyclin T1) triggers elongation. FUS, an EWSR1 homologue, has been reported to regulate RNAPII Ser2 phosphorylation¹¹. Therefore, the dysregulated transcriptional response of Ewing sarcoma could be due to EWS–FLI1 interfering with wild-type EWSR1 in regulating transcription. In an *in vitro* kinase assay using purified recombinant proteins, EWSR1 inhibited phosphorylation of the RNAPII CTD by CDK9 (Fig. 1d, Extended Data Fig. 2a) whereas EWS–FLI1 did not (Extended Data Fig. 2b). EWSR1 depletion in U2OS cells increased RNAPII phosphorylation, confirming the results of the kinase assay (Fig. 1e). Immunoblotting of Ewing sarcoma cell lysates indicated high levels of phospho-Ser2/Ser5 RNAPII compared to IMR90 cells (Fig. 1f), and depletion of EWS–FLI1 in TC32 cells significantly decreased RNAPII phosphorylation (Fig. 1g). Notably, wild-type EWSR1 levels were not affected by EWS–FLI1 knockdown. These data suggest that EWS–FLI1 increased basal levels of transcription, either directly, or indirectly by interfering with EWSR1 activity. Accordingly, EWSR1-depleted cells and Ewing sarcoma cell lines and tumours were more sensitive to blockade of transcription by camptothecin (a topoisomerase I inhibitor) (Extended Data Fig. 2c, d).

DNA damage induces global suppression of transcription (involving BRCA1¹²) followed by gradual recovery. We therefore evaluated the transcription response and recovery following etoposide exposure using incorporation of ethynyl uridine (EU) into RNA (Fig. 1h). Unlike IMR90 cells, which displayed a characteristic decrease in EU incorporation two hours post damage followed by recovery, TC32 Ewing

¹Department of Cell Systems and Anatomy, University of Texas Health at San Antonio, San Antonio, Texas 78229, USA. ²Greehey Children's Cancer Research Institute, University of Texas Health at San Antonio, San Antonio, Texas 78229, USA. ³Department of Pediatric Oncology, Dana-Farber Cancer Institute and Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts 02215, USA. ⁴The Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. ⁵Departments of Oncology and Pediatrics, Georgetown University, Washington DC 20057, USA. ⁶Departments of Pediatrics and Pathology, University of Michigan, Ann Arbor, Michigan 48109, USA. ⁷Center for Childhood Cancer and Blood Diseases, Nationwide Children's Hospital, Columbus, Ohio 43205, USA. ⁸Mays Cancer Center, University of Texas Health at San Antonio, Texas 78229, USA. ⁹Department of Epidemiology and Biostatistics, University of Texas Health at San Antonio, San Antonio, Texas 78229, USA.

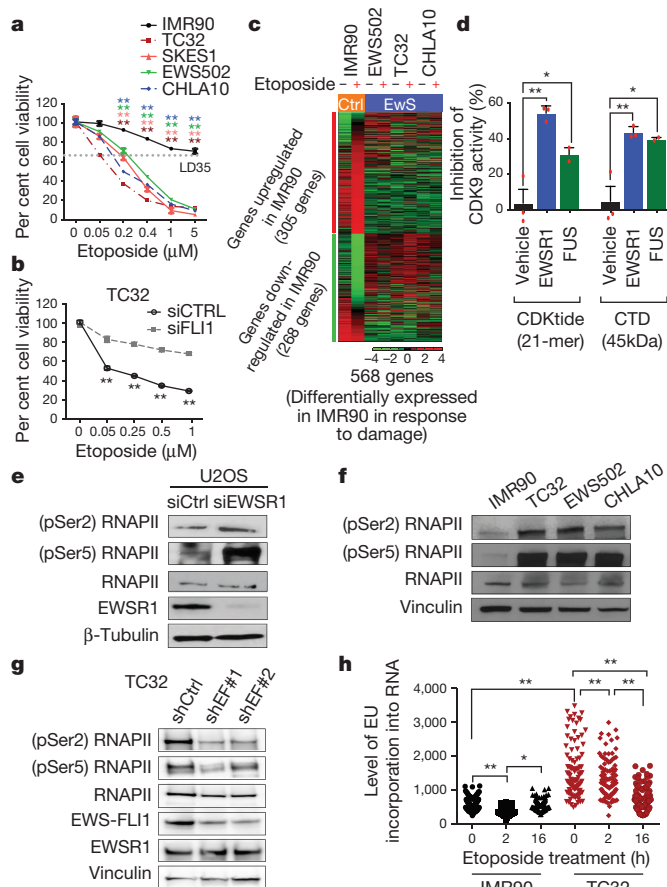


Figure 1 | Ewing sarcoma dysregulates transcription in response to damage. **a**, Cell viability following etoposide treatment. Etoposide dose causing 35% lethality (LD35, dotted grey line) was used for further experiments. Mean \pm s.d., $n = 4$ technical replicates, one-way ANOVA. **b**, Etoposide-induced TC32 cytotoxicity after EWS-FLI1 knockdown (siFLI1). $n = 4$ transfection replicates, two-tailed t -tests. **c**, Heat map of damage-induced differential gene expression. **d**, CDK9 kinase activity inhibition by recombinant EWSR1 and FUS proteins on CDKtide or CTD substrates. $n = 3$ technical replicates, one-way ANOVA. **e–g**, Levels of phosphorylated Ser2/Ser5 RNAPII in U2OS cells with EWSR1 knockdown (**e**), IMR90 cells versus TC32 cells (**f**), and TC32 cells with EWS-FLI1 knockdown (**g**). **h**, Transcriptional activity after etoposide treatment. Centre at median, $n = 100$ cells, two-way ANOVA. Mean \pm s.e.m., * $P < 0.05$, ** $P < 0.005$.

sarcoma cells showed a significantly higher basal transcription level, similar to that seen in EWSR1-depleted cells (Extended Data Fig. 2e), and a delayed decrease in transcription.

Alterations in regulation of transcription could result in the accumulation of R-loops (three-stranded nucleic acid structures comprising a DNA–RNA hybrid and non-template single-stranded DNA)¹³. Several pieces of evidence advocate a potential role for EWSR1 and EWS-FLI1 in regulating R-loop accumulation: both proteins regulate RNAPII elongation⁹ and interact with splicing machinery^{9,14,15}, each of which is conducive to R-loop formation¹⁶. Using the RNA–DNA hybrid-specific S9.6 antibody to probe genomic DNA, we discovered that Ewing sarcoma cell lines displayed nearly fourfold higher levels of R-loops compared to IMR90 cells (Fig. 2a). RNaseH treatment (Fig. 2a) or RNaseH1 expression (Extended Data Fig. 3a) substantially decreased the R-loop signal. R-loop accumulation was induced by EWSR1 depletion (Fig. 2b) or by EWS-FLI1, as shown by expression in U2OS cells (Fig. 2b) or knockdown in TC32 cells (Extended Data Fig. 3b). A DNA-binding mutant EWS-FLI1 (Extended Data Fig. 3c) also induced accrual of R-loops, suggesting that the N-terminal protein interaction domain that is common to EWSR1 and EWS-FLI1 is important in promoting

R-loop accumulation. Notably, Ewing sarcoma cells did not display a damage-induced reduction in R-loops to the same extent as IMR90 cells (Extended Data Fig. 3d, e), consistent with the EU-incorporation data (Fig. 1h). These results were corroborated by immunofluorescence analysis of nucleoplasmic R-loops (Fig. 2c, Extended Data Fig. 3f).

To further characterize the R-loops present in Ewing sarcoma, we conducted high-throughput sequencing and analysis of genomic DNA–RNA hybrids by immunoprecipitation (DRIP-seq)¹⁷ in Ewing sarcoma and control cell lines with or without damage. RNaseH-treated DNA served as a negative control. We found extensive R-loops throughout the genome in EWSR1-depleted cells compared to control cells (by read coverage and depth), and this difference was greater in Ewing sarcoma cells (Fig. 2d, Extended Data Fig. 4a). Comparison of DRIP-seq and RNA sequencing (RNA-seq) data indicated a predominance of R-loops in regions containing highly expressed genes (Extended Data Fig. 4b). We examined the co-occurrence of reported EWS-FLI1 binding sites¹⁸ with R-loops and found very strong enrichment (Extended Data Fig. 4c), especially at highly expressed genes (top 16%). We also observed an increased propensity for RNAPII binding (by RNAPII chromatin immunoprecipitation and sequencing, ChIP-seq) in the same regions as R-loops compared to surrounding genomic regions (Extended Data Fig. 4c), which was confirmed at well-established R-loop sites (Extended Data Fig. 4d).

Unresolved R-loops are deleterious to the cell, as they potentially block replication machinery progression and result in stalled or collapsed replication forks^{13,16}. Analysis of well-known markers of replication stress indicated elevated basal levels of activated (phosphorylated) ATR, CHK1 and RPA2 in Ewing sarcoma cells (Fig. 2e). Significant sensitivity to ATR inhibition was observed not only in Ewing sarcoma cells (as previously reported¹⁹), but also in EWSR1-depleted U2OS cells (Fig. 2f). RNaseH1 overexpression suppressed ATR pathway activation (Extended Data Fig. 5a) and increased the rate of cell proliferation (Fig. 2g), confirming that replication stress in Ewing sarcoma is due to accumulated R-loops.

R-loop accumulation is generally associated with increased DNA damage and homologous recombination¹³. Ewing sarcoma cells exhibit high levels of DNA damage²⁰ (measured by the P53-binding protein TP53BP1 foci; Fig. 3a, b) compared to IMR90 cells. However, there was an absence of ionizing radiation-induced RAD51 foci (Fig. 3a, c). Basal levels of RAD51 foci were higher in Ewing sarcoma cells than in IMR90 cells (Fig. 3c), although this may reflect increased replication stress²¹. We used the direct-repeat GFP assay (DR-GFP; Extended Data Fig. 5b) integrated into U2OS cells to evaluate endonuclease-induced double strand break (DSB) repair by homologous recombination²². Expression of either EWS-FLI1 or EWS-ERG (the second most common Ewing sarcoma translocation²³) significantly reduced homologous recombination capacity (Fig. 3d). As EWS-FLI1 binds EWSR1 through their shared N-terminal domain⁶, we investigated whether the suppression of homologous recombination was due to a loss of EWSR1 function. As suspected, either expression of the EWSR1 N-terminal domain alone (Fig. 3d) or expression of two independent small inhibitory RNAs (siRNAs) against EWSR1 (Fig. 3e) also reduced homologous recombination frequency.

Given the similarity in phenotypes between Ewing sarcoma and BRCA1/2 mutant breast cancer (Extended Data Table 1c), we investigated whether BRCA1 function was altered in Ewing sarcoma. It is noteworthy that our gene expression analysis, comparing Ewing sarcoma and control cells, identified significant enrichment for a BRCA1-mutated gene set (Extended Data Fig. 1g). Ewing sarcoma cells have robust BRCA1 expression (Extended Data Fig. 5c) with no known mutations. To test a functional absence, we overexpressed BRCA1 and found restored homologous recombination in the context of EWS-FLI1 expression (Fig. 3f), suggesting a functional impairment, but unexpectedly did not do so when EWSR1 was depleted (Fig. 3g).

PARP1 inhibition is synthetic lethal with BRCA1 deficiency in replicating cells^{24,25} owing to the absence of homologous recombination.

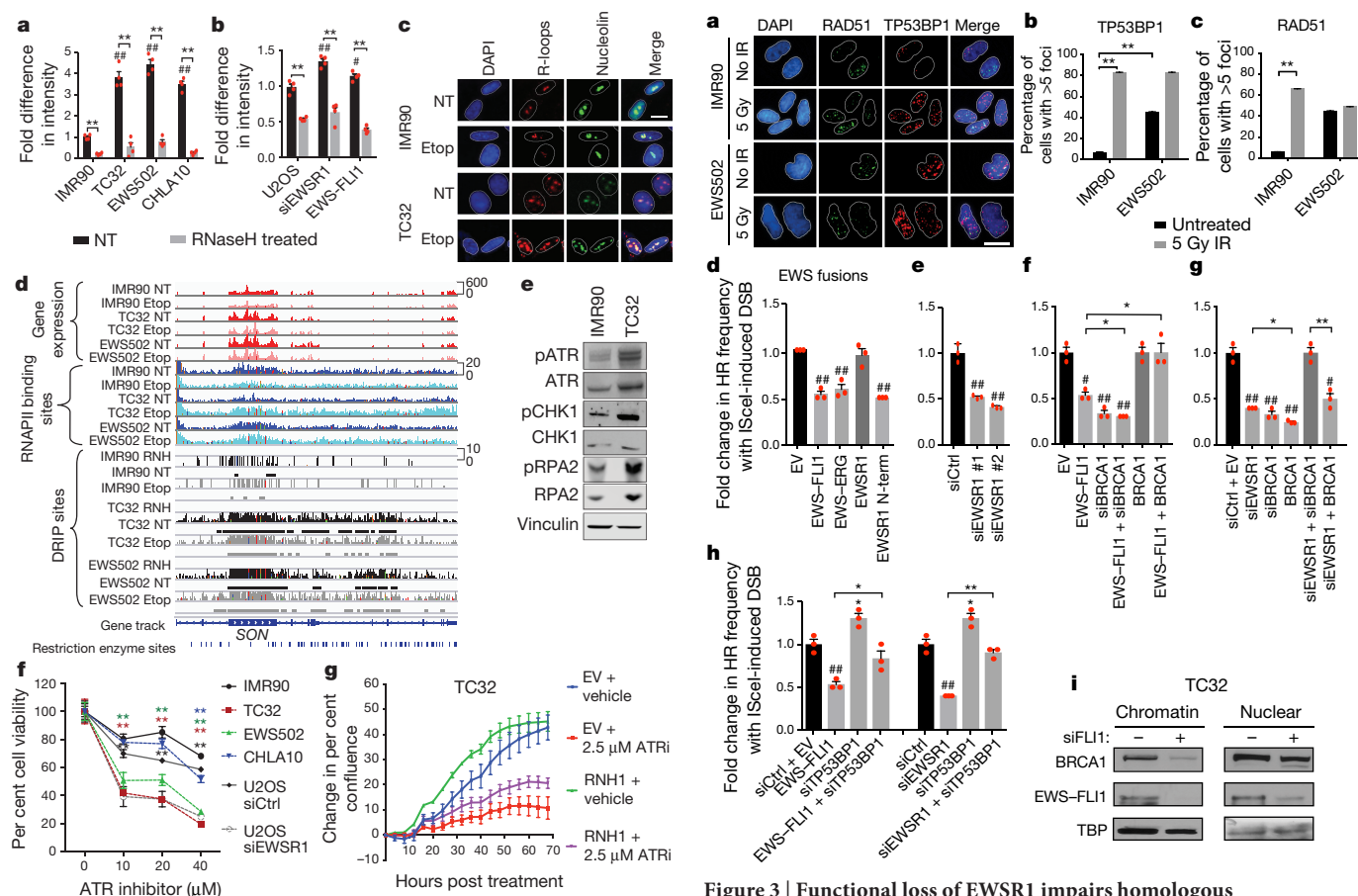


Figure 2 | R-loop accumulation in Ewing sarcoma. **a, b**, Fold-difference in genomic R-loops in IMR90 versus Ewing sarcoma cells (**a**) and U2OS cells with indicated transfections (**b**); *n* = 4 technical replicates. **c**, Representative immunofluorescence images of nuclei (DAPI), R-loops (S9.6) and nucleoli (nucleolin). Scale bar, 25 μ m. **d**, Thirty-five-kilobase region surrounding the SON gene containing RNA-seq (red), RNAPII ChIP-seq (blue), DRIP-seq (black) and RNaseH-treated (RNH) tracks. Etos, etoposide-treated. Track height represents read counts. **e**, Immunoblots of indicated replication stress proteins. **f**, Cell viability response to ATR inhibitor; *n* = 4 technical replicates. **g**, Proliferation rate of TC32 cells overexpressing RNaseH1 (RNH1) after ATR inhibition (ATRi). EV, empty vector; *n* = 3 transfection replicates. Mean \pm s.e.m., one-way ANOVA compared to IMR90 control, two-tailed *t*-test within each cell line. #, **P* < 0.05; ##, ***P* < 0.005, # shows significance compared to control (NT).

Thus, impaired BRCA1 function could provide a molecular basis for the high sensitivity of Ewing sarcoma to PARP1 inhibitors such as Olaparib²⁰ (Extended Data Fig. 6a–c) and etoposide-induced DNA breaks. Mutation of TP53BP1 circumvents the need for BRCA1 in homologous recombination, partially restoring homologous recombination and conferring some resistance to chemotherapy²⁶. Depletion of TP53BP1 restored homologous recombination in the presence of either EWSR1 knockdown or EWS-FLI1 expression (Fig. 3h), consistent with a functional deficiency of BRCA1. Knockdown of TP53BP1 also moderately increased resistance to etoposide in Ewing sarcoma cells (Extended Data Fig. 6e). Together, our data indicate that Ewing sarcoma phenocopies BRCA1-deficient tumours and suggest that secondary mutations in TP53BP1 are a potential chemoresistance mechanism.

BRCA1 has been shown to associate with the elongating transcription complex²⁷ and with R-loops²⁸. Sequestration of BRCA1 with transcription complexes could prevent its redistribution to exogenous damage and therefore explain its functional absence in Ewing sarcoma. While overall levels of BRCA1 protein were similar between Ewing

Figure 3 | Functional loss of EWSR1 impairs homologous recombination. **a**, Representative immunofluorescence images of TP53BP1 and RAD51 foci. Scale bar, 25 μ m. **b, c**, Number of cells with more than five TP53BP1 (**b**) or RAD51 (**c**) foci. *n* = 85 cells, two-tailed *t*-test. **d–h**, Homologous recombination (HR) frequency with indicated transfections (efficiency demonstrated in Extended Data Figs 5d, 6d). *n* = 3 transfection replicates, one-way ANOVA. **i**, Immunoblots for BRCA1 and EWS-FLI1 in subcellular fractions of TC32 cells with EWS-FLI1 knockdown. Mean \pm s.e.m.; #, **P* < 0.05; ##, ***P* < 0.005, # shows significance relative to control (black bar).

sarcoma and control cells (Extended Data Fig. 6f), comparison of the subcellular fractions of U2OS versus TC32 cells revealed redistribution of BRCA1. Chromatin-bound BRCA1 was particularly enriched in Ewing sarcoma (Fig. 3i), but was substantially reduced by EWS-FLI1 knockdown. EWSR1 depletion did not result in a similar sequestration of BRCA1 in the chromatin fraction, again highlighting the difference between EWSR1 loss and EWS-FLI1 expression with regard to BRCA1 function (Extended Data Fig. 6g).

BRCA1 preferentially associates with phospho-RNAPII in undamaged cells²⁷. Considering the increased amount of phospho-RNAPII in Ewing sarcoma cells, we examined the interaction between endogenous phospho-RNAPII and BRCA1 by co-immunoprecipitation (Fig. 4a, Extended Data Fig. 7a) in nuclear lysates. An increased proportion of phospho-RNAPII was immunoprecipitated by BRCA1 in Ewing sarcoma, highlighting the redistribution of BRCA1 to transcription complexes. This interaction did not diminish following damage as expected²⁷ and seen in control cell lines. We also confirmed the lack of interaction between BRCA1 and unphosphorylated RNAPII. Subsequent to release from the transcription machinery, the association of BRCA1 with the retinoblastoma binding protein 8 (RBBP8 or CtIP) increases to promote removal of TP53BP1²⁹ and DSB repair by homologous recombination. An increase in damage-induced interaction between BRCA1 and CtIP was observed in control but not Ewing sarcoma cells. Confirming the functional impairment of BRCA1 in DSB repair, we performed BRCA1 ChIP in U2OS cells after

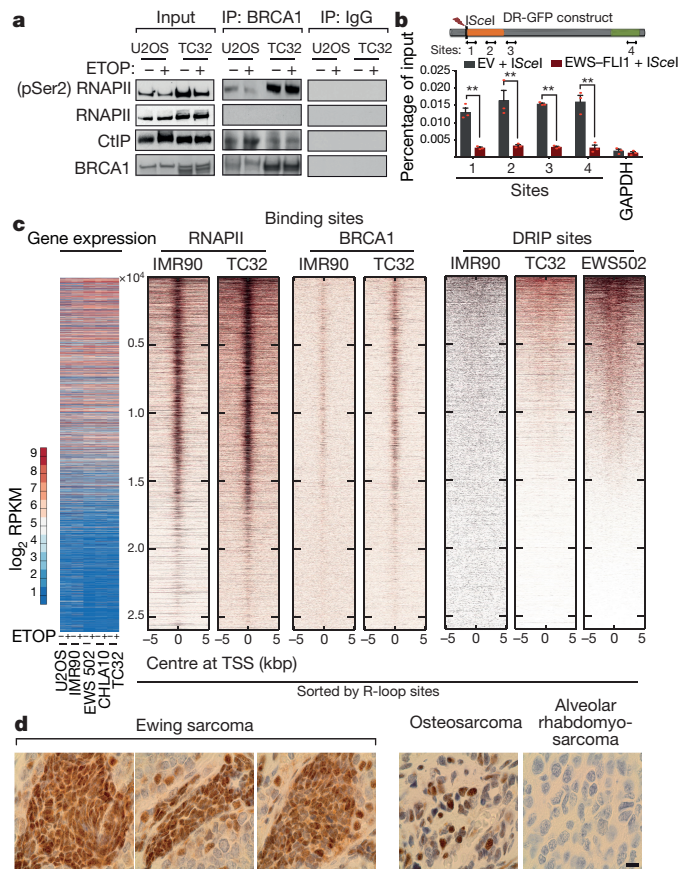


Figure 4 | BRCA1 is retained at transcriptional complexes in Ewing sarcoma. **a**, Immunoblots of indicated proteins after BRCA1 immunoprecipitation. **b**, Schematic of qPCR sites within DR-GFP locus relative to break site used to measure BRCA1 occupancy after EWS-FLI1 overexpression (graph below); GAPDH serves as negative control. Mean \pm s.e.m., $n = 3$ technical replicates; $**P < 0.005$, two-tailed t -test. **c**, Whole-genome heat maps representing correlation between gene expression, RNAPII, BRCA1 binding sites, and DRIP loci ordered by DRIPs, centred on TSS. **d**, Representative immunohistochemistry images showing R-loop staining across a sarcoma tissue microarray. Scale bar, 10 μ m.

endonuclease-induced DSB at the DR-GFP locus and observed a loss of BRCA1 recruitment in cells transfected with EWS-FLI1 compared to vector-transfected controls (Fig. 4b).

Finally, we performed an in-depth analysis of the association of BRCA1 with R-loops in control and Ewing sarcoma cell lines with and without exposure to damage (BRCA1 ChIP-seq, Fig. 4c, Extended Data Fig. 8a). BRCA1 binding (Extended Data Fig. 7b, c) decreased significantly upon damage in control cells, but not as much in Ewing sarcoma cells. We detected BRCA1 binding at a well-known R-loop region¹⁷ (Extended Data Fig. 7d, e), with a significant damage-dependent decrease in controls but not in Ewing sarcoma cell lines. Conversely, we also confirmed the presence of R-loops at BRCA1 binding sites (Extended Data Fig. 7f). Genome-wide maps of BRCA1 sites centred on transcription start sites (TSS) indicated that highly expressed gene loci were associated with both BRCA1 and RNAPII localization (Fig. 4c, Extended Data Fig. 8b). There was strong enrichment for BRCA1 and RNAPII binding at R-loops, particularly in the Ewing sarcoma cells (Extended Data Fig. 8c). We found increased co-localization between BRCA1- and RNAPII-bound TSS in Ewing sarcoma (2,569 genes) compared to IMR90 cells (269 genes) (Extended Data Fig. 9a–d), corroborating the co-immunoprecipitation data. We also observed an increase in enrichment (peak height) of these two proteins in TC32 compared to IMR90 cells. Collectively, the above results indicate that in Ewing sarcoma, BRCA1 is retained at stalled

transcription complexes associated with R-loops, presumably to mitigate damage associated damage.

In summary, our work provides a detailed examination of basal and damage-responsive R-loops in Ewing sarcoma. We delineate the basis of EWS-FLI1-mediated chemosensitivity and conclude that interference with wild-type EWSR1 is responsible for a large part of EWS-FLI1 function. We confirmed the prevalence of accumulated R-loops in primary Ewing sarcoma tumours compared to other sarcomas by immunohistochemical analysis on a sarcoma tissue microarray stained with S9.6 antibody (Fig. 4d, Extended Data Fig. 10). Mutations in EWSR1 and its homologues are associated with several therapeutically challenging cancers³⁰. It is tempting to speculate that the causal transcription stress phenotypes may extend to these tumours. We envision the use of agents that induce transcription or replication stress as potentially effective augmentative treatment strategies in various tumours³⁰ associated with EWSR1 translocations.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 September 2015; accepted 11 January 2018.

Published online 7 March 2018.

1. Turc-Carel, C., Philip, I., Berger, M. P., Philip, T. & Lenoir, G. M. Chromosome study of Ewing's sarcoma (ES) cell lines. Consistency of a reciprocal translocation t(11;22)(q24;q12). *Cancer Genet. Cytogenet.* **12**, 1–19 (1984).
2. Paronetto, M. P. Ewing sarcoma protein: a key player in human cancer. *Int. J. Cell Biol.* **2013**, 642853 (2013).
3. Li, H. *et al.* Ewing sarcoma gene EWS is essential for meiosis and B lymphocyte development. *J. Clin. Invest.* **117**, 1314–1323 (2007).
4. Yang, W. *et al.* Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* **41**, D955–D961 (2013).
5. Kim, N. & Jinks-Robertson, S. Transcription as a source of genome instability. *Nat. Rev. Genet.* **13**, 204–214 (2012).
6. Spahn, L. *et al.* Homotypic and heterotypic interactions of EWS, FLI1 and their oncogenic fusion protein. *Oncogene* **22**, 6819–6829 (2003).
7. Embree, L. J., Azuma, M. & Hickstein, D. D. Ewing sarcoma fusion protein EWSR1/FLI1 interacts with EWSR1 leading to mitotic defects in zebrafish embryos and human cell lines. *Cancer Res.* **69**, 4363–4371 (2009).
8. Bertolotti, A. *et al.* EWS, but not EWS-FLI-1, is associated with both TFIIID and RNA polymerase II: interactions between two members of the TET family, EWS and hTAFII68, and subunits of TFIIID and RNA polymerase II complexes. *Mol. Cell Biol.* **18**, 1489–1497 (1998).
9. Yang, L., Chansky, H. A. & Hickstein, D. D. EWS-FLI-1 fusion protein interacts with hyperphosphorylated RNA polymerase II and interferes with serine-arginine protein-mediated RNA splicing. *J. Biol. Chem.* **275**, 37612–37618 (2000).
10. Phatnani, H. P. & Greenleaf, A. L. Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev.* **20**, 2922–2936 (2006).
11. Schwartz, J. C. *et al.* FUS binds the CTD of RNA polymerase II and regulates its phosphorylation at Ser2. *Genes Dev.* **26**, 2690–2695 (2012).
12. Hill, S. J. *et al.* Systematic screening reveals a role for BRCA1 in the response to transcription-associated DNA damage. *Genes Dev.* **28**, 1957–1975 (2014).
13. Aguilera, A. & García-Muse, T. R loops: from transcription byproducts to threats to genome stability. *Mol. Cell* **46**, 115–124 (2012).
14. Paronetto, M. P., Miñana, B. & Valcárcel, J. The Ewing sarcoma protein regulates DNA damage-induced alternative splicing. *Mol. Cell* **43**, 353–368 (2011).
15. Selvanathan, S. P. *et al.* Oncogenic fusion protein EWS-FLI1 is a network hub that regulates alternative splicing. *Proc. Natl Acad. Sci. USA* **112**, E1307–E1316 (2015).
16. Hamperl, S. & Cimprich, K. A. The contribution of co-transcriptional RNA:DNA hybrid structures to DNA damage and genome instability. *DNA Repair (Amst.)* **19**, 84–94 (2014).
17. Ginno, P. A., Lott, P. L., Christensen, H. C., Korf, I. & Chédin, F. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol. Cell* **45**, 814–825 (2012).
18. Riggi, N. *et al.* EWS-FLI1 utilizes divergent chromatin remodeling mechanisms to directly activate or repress enhancer elements in Ewing sarcoma. *Cancer Cell* **26**, 668–681 (2014).
19. Nieto-Soler, M. *et al.* Efficacy of ATR inhibitors as single agents in Ewing sarcoma. *Oncotarget* **7**, 58759–58767 (2016).
20. Brenner, J. C. *et al.* PARP-1 inhibition as a targeted strategy to treat Ewing's sarcoma. *Cancer Res.* **72**, 1608–1613 (2012).
21. Zellweger, R. *et al.* Rad51-mediated replication fork reversal is a global response to genotoxic treatments in human cells. *J. Cell Biol.* **208**, 563–579 (2015).
22. Pierce, A. J., Johnson, R. D., Thompson, L. H. & Jasin, M. XRCC3 promotes homology-directed repair of DNA damage in mammalian cells. *Genes Dev.* **13**, 2633–2638 (1999).

23. Sankar, S. & Lessnick, S. L. Promiscuous partnerships in Ewing's sarcoma. *Cancer Genet.* **204**, 351–365 (2011).
24. Bryant, H. E. *et al.* Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* **434**, 913–917 (2005).
25. Farmer, H. *et al.* Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434**, 917–921 (2005).
26. Bouwman, P. *et al.* 53BP1 loss rescues BRCA1 deficiency and is associated with triple-negative and BRCA-mutated breast cancers. *Nat. Struct. Mol. Biol.* **17**, 688–695 (2010).
27. Krum, S. A., Miranda, G. A., Lin, C. & Lane, T. F. BRCA1 associates with processive RNA polymerase II. *J. Biol. Chem.* **278**, 52012–52020 (2003).
28. Hatchi, E. *et al.* BRCA1 recruitment to transcriptional pause sites is required for R-loop-driven DNA damage repair. *Mol. Cell* **57**, 636–647 (2015).
29. Escribano-Díaz, C. *et al.* A cell cycle-dependent regulatory circuit composed of 53BP1-RIF1 and BRCA1-CtIP controls DNA repair pathway choice. *Mol. Cell* **49**, 872–883 (2013).
30. Riggi, N., Cironi, L., Suvà, M. L. & Stamenkovic, I. Sarcomas: genetics, signalling, and cellular origins. Part 1: the fellowship of TET. *J. Pathol.* **213**, 4–20 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank F. Chedin for the DRIP-seq protocol and genome-wide restriction enzyme sites file, and R. Crouch, J. Stark and Y. Shiao for plasmids. We are grateful to the UTH-SA/Cancer Center Sequencing core and the Histology & Immunohistochemistry Core at UTH-SA. This work was funded by the NIH (K22ES012264, 1R15ES019128, 1R01CA152063), a Voelcker

Fund Young Investigator Award and CPRIT (RP150445) to A.J.R.B.; CPRIT (RP101491), a Translational Science Training Across Disciplines Scholarship (UTHSA) and an NCI postdoctoral training grant (T32CA148724) to A.G.; CPRIT (RP150445) to J.C.R.; NIH (P30CA054174) to Mays Cancer Center; NCI (R01CA204915) and Curing Kids Cancer to K.S.; NIH CTSA (1UL1RR025767-01, P30CA054174) and CPRIT (RP120685-C2) to Y.C.; NIH (1R01CA140394) to S.L.L. and NIH (1R01CA134605) to E.R.L.

Author Contributions A.G. and A.J.R.B. conceived and designed the study and wrote the manuscript. A.G. conducted the majority of the research. J.C.R. contributed to replication of experiments for homologous recombination and BRCA1 localization. E.G. performed EU incorporation experiments. L.A.L. performed RNaseH rescue of replication stress and EU incorporation experiments. E.L., L.C., V.P.M. and X.B. provided technical support. A.B.I. and K.S. performed EWS-FLI1 knockdown for the phospho-RNAPII experiment. S.R. and J.A.T. provided recombinant EWS-FLI1 protein. Y.C. conducted bioinformatics analysis support. S.L.L. and E.R.L. provided reagents/insights.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to A.J.R.B. (bishopa@uthscsa.edu).

Reviewer Information *Nature* thanks B. Braun and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Cell culture and transfections. Ewing sarcoma cell lines (TC32, CHLA10 and CHLA258) were obtained from Children's Oncology Group; EWS502 was obtained from S. Lessnick; and SKES1 and RDES were obtained from S. Mooberry. As controls, IMR90 (CCL-186, primary human fetal fibroblast) and U2OS (HTB-96, paediatric osteosarcoma) cell lines were obtained from the American Type Culture Collection (ATCC). IMR90 and U2OS cells were cultured in DMEM (Corning); TC32, EWS502 and RDES cells in RPMI-40 (Corning); SKES1 cells in McCoy's (Corning); and CHLA10 and CHLA258 cells in IMDM (HyClone); all cultured media were supplemented with 10% heat-inactivated fetal bovine serum (Atlanta Biologicals) and 1% antibiotic/antimycotic solution (Corning). *Drosophila melanogaster* Kc167 cells were purchased from the *Drosophila* Genomics Resource Center and maintained in Schneider medium (Invitrogen) supplemented with 10% heat-inactivated fetal bovine serum at 22°C in a humidified chamber. Human cells were maintained at 37°C in a humidified atmosphere with 5% CO₂ and tested for mycoplasma contamination. Cell lines were procured from reliable sources and additionally TC32, CHLA10, U2OS and DR-GFP U2OS cell lines were authenticated by STR analysis.

All transfections were carried out using Lipofectamine RNAiMax or Lipofectamine 3000 (Invitrogen) following the manufacturer's instructions. For RNaseH1 transfections, 3×10^6 TC32 cells were transfected with 13 µg of plasmid DNA using Amaxa nucleofection (program X-001). Gene knockdowns were performed by reverse transfection whereas plasmid transfections were performed 24 h after seeding. The siRNAs used in this study include: BRCA1 and EWSR1 (Life Technologies), EWSR1, BRCA1, FLI1 and TP53BP1 (Santa Cruz). For some experiments, lentiviral transduction of TC32 cells with control short hairpin RNA (shCtrl) (clone RHZ4743, Life Technologies) or shFLI1 (clone V2THS227524 or V3THS414176, Life Technologies) was performed. EWS-FLI1, EWS-FLI1 R2L2, EWS-ERG, EWSR1 (full length) and shEWSR1 in pMSCV vector were obtained from S. Lessnick. The EWSR1 N-terminal domain in pLX304 vector was purchased from DNASU. GFP-RNaseH1 and FLAG-RNaseH1 plasmids³¹ were a gift from R. Crouch (Eunice Kennedy Shriver National Institute of Child Health and Human Development). HA-BRCA1 was obtained from Y. Shiio (UTH-SA). IScel in pCAGGS vector was a gift from M. Jasin (Memorial Sloan Kettering Cancer Center) and J. Stark (City of Hope Cancer Center). All siRNA and plasmid transfections were accompanied with control siRNA or empty vector respectively.

Cell viability. Cells were seeded at 30% confluence in 96- or 384-well plates with or without reverse transfection with siRNA. For plasmid expression, cells were transfected in 60-mm dishes first and then split into 96- or 384-well plates. Cells were treated with different drugs or inhibitors on the next day, and cell viability was evaluated after 48–72 h using Celltiter-Glo (Promega). An etoposide dose causing 35% cell death (LD35) was used for most experiments unless indicated otherwise. Each condition was tested at least in triplicate. Chemicals used in cell viability experiments include etoposide (E1383, Sigma), camptothecin (C9911, Sigma), VE-821 (ATR inhibitor, Selleck Chemicals) and Olaparib (PARP1 inhibitor, Selleck Chemicals). Drug screening data from an independently published study⁴ were obtained from <http://www.cancerrxgene.org/>.

Immunoblotting and immunoprecipitation. Whole-cell lysates were prepared using RIPA buffer according to standard protocols. A subcellular protein fractionation kit for cultured cells (Thermo Fisher Scientific) was used to extract nuclear and chromatin fractions. Cell lysates were separated on either precast 3–8% or 4–12% gradient gels (Invitrogen) or laboratory-prepared gels and transferred onto nitrocellulose membrane. All blots were incubated with primary antibodies overnight and developed using enhanced chemiluminescence (Super ECL, Thermo Fisher Scientific). Antibodies used in this study include FLI1 (ab15289, Abcam), EWSR1 (ab133288 and ab54708, Abcam), RNAPII phospho Ser2 (MMS-129R-200, Covance), RNAPII phospho Ser5 (61085, Active Motif), RNAPII (ab817, Abcam), RNA–DNA hybrids (S9.6, ENH001, Kerafast), single-stranded DNA (ssDNA, MAB3034, Millipore), RNaseH1 (15606-1-AP, Proteintech), nucleolin (sc-13057, Santa Cruz), TP53BP1 (A300-272A, Bethyl labs), RAD51 (70-005-EX, Cosmo), ATR phospho Ser428 (cs28539, Cell Signaling), ATR (sc-1887, Santa Cruz), CHK1 phospho Ser317 (cs2344, Cell Signaling), CHK1 (cs2345, Cell Signaling), RPA2 phospho Ser33 (A300-246A, Bethyl labs), RPA2 (ab2175, Abcam), BRCA1 (sc642, Santa Cruz and 07-434, Millipore), CtIP (cs9201, Cell Signaling), Flag-tag (ab1162, Abcam), Tata-binding protein (TBP, ab818, Abcam), β-actin (ab16039, Abcam), β-tubulin (cs2128, Cell Signaling), vinculin (cs13901, Cell Signaling), GAPDH (cs5174, Cell Signaling), lamin B1 (cs9087, Cell Signaling), Sp1 (sc-59, Santa Cruz), histone H3 (cs9715, Cell Signaling) and secondary antibodies goat anti-mouse IgG-HRP (sc-2060, Santa Cruz), goat anti-rabbit IgG-HRP (sc-2030, Santa Cruz) and goat anti-rat IgG-HRP (sc-2065, Santa Cruz). Western blot experiments were repeated with independent sample preparations three to five times.

All co-immunoprecipitation experiments were done with endogenous proteins. In brief, cells from nearly confluent 15-cm plates treated with either vehicle or

an LD₅₀ dose of etoposide for 2 h were collected and lysed in cytosolic extraction buffer (low salt buffer: 20 mM Hepes, pH 7.4, 0.5% Nonidet P-40, 10 mM NaCl supplemented with Halt protease and phosphatase inhibitors) for 30 min on ice. The nuclei were subsequently extracted in a high salt buffer (20 mM Hepes, 0.5% Nonidet P-40, 1.5 mM MgCl₂, 0.5 M NaCl) for 45 min on a rocker at 4°C. The lysates were then diluted to physiological salt concentration (150 mM NaCl) and incubated for another 45 min. The nuclear lysates (0.5–0.75 mg) were pre-cleared using protein A/G beads and incubated overnight with 2 µg antibody on a rocker at 4°C. Settled equilibrated protein A/G beads (25 µl) (or protein L for IgM antibody) were then added to the antibody-lysate mixture and incubated for further 6 h. Bound complexes were washed three times in lysis buffer before elution with Invitrogen NuPage loading buffer. Eluted proteins were evaluated by immunoblotting and compared to inputs (10% of the amount used for immunoprecipitation). Co-immunoprecipitation experiments were repeated with biological replicates at least three times in independent sample preparations.

In vitro RNAPII phosphorylation assay. The following purified recombinant proteins were purchased: RNAPII CTD fragment (POLR2A-1149H, Creative BioMart), EWSR1 (TP303709, Origene), FUS (TP301808, Origene) and CDK9/cyclin T1 (PV4131, Thermo Fisher Scientific). The *in vitro* kinase assay was performed using the Adapta kinase assay (PV5099, Thermo Fisher Scientific) according to the manufacturer's instructions. The template used for kinase activity was either 50 µM CDK7/9tide (triheptad repeat peptide) provided in the kit (PV5090, Thermo Fisher Scientific), or 12 ng RNAPII CTD fragment. EWSR1 or FUS (1 µM) was added to the substrate and 10 µM ATP kinase buffer and pre-incubated for 15 min. CDK9/cyclin T1 (1.77 µg/ml) was then added and the kinase reaction was allowed to proceed for a further 45 min. ADP and ATP (10 µM each) served as positive and negative controls, respectively. Finally, EU-labelled antibody and Alexa Fluor-647-labelled tracer were added and the level of ATP consumption was measured 30 min later using BMG Labtech Pherastar microplate reader. Each condition was tested with technical quadruplicates and the overall experiment was repeated at least twice for independent validation.

Transcription recovery assay. Cells grown on fibronectin-coated coverslips were treated with LD₅₀ doses of etoposide for 0, 2 or 16 h, followed by incubation in medium containing 0.5 mM 5-ethynyl uridine (EU, Thermo Fisher Scientific) for 30 min. After incubation, cells were fixed, permeabilized and subjected to Click-iT RNA reaction (Click-iT RNA Alexa Flour 598 Imaging kit, Thermo Fisher Scientific) according to the manufacturer's instructions. Washed coverslips were mounted onto glass slides using ProLong Antifade Mounting Solution with DAPI (Thermo Fisher Scientific) and imaged using a Zeiss microscope at 63×. A minimum of 100 cells was analysed for each condition. Image analysis was performed using ImageJ. The overall experiment was independently verified three times.

Dot blot for R-loops. Restriction enzyme-digested genomic DNA (0.5 µg) was loaded on to pre-wet H⁺ nylon membrane. The membrane was washed twice with dH₂O, rinsed in 2× SSC buffer and then left to air dry at room temperature. For ssDNA, an additional denaturation step (incubation in 0.5 N NaOH, 1.5 M HCl for 10 min), followed by a 10 min incubation in neutralization buffer (1 M NaCl, 0.5 M Tris-HCl pH 7) was performed. The membrane was then blocked with 1× TBS containing 5% non-fat dry milk and incubated with the primary antibody (R-loops S9.6 antibody and ssDNA) in blocking buffer overnight. The blots were analysed using ImageJ software or LI-COR Image Studio to measure signal intensity from genomic DNA, RNaseH-treated DNA and ssDNA. ssDNA signal was used to normalize R-loop signal. Dot blot experiments were performed with technical quadruplicates and repeated at least twice with independent sample preparation for validation.

Immunofluorescence. Cells were seeded on fibronectin-coated coverslips. Following knockdown and/or 6-h etoposide treatment, cells were fixed with 4% paraformaldehyde. For R-loops, a pre-extraction step was carried out by treating the cells with 0.1% Triton-X 100 in PBS for 30 s at room temperature. After permeabilization with Triton-X 100 for 10 min, cells were blocked for 1 h with 1% BSA, 4% goat serum followed by overnight incubation with primary antibodies in blocking buffer. Coverslips were then incubated with Alex Fluor 488/568 conjugated secondary antibodies (Life Technologies). The cells were then stained with DAPI and coverslips were mounted on slides using Vectashield (Vector Labs). Cells were imaged using a Zeiss microscope at 40× or 63×. At least 100 nuclei per dataset were sampled and a minimum of 80 nuclei per condition was used for quantification of immunofluorescence intensity or foci counting. Image analysis was done using Adobe Photoshop software. Immunofluorescence experiments were repeated for validation.

Immunohistochemistry. The sarcoma tissue microarray (T264) was purchased from US Biomax. Tissue microarrays were treated with 1 mM EDTA pH 8 for 40 min at 95°C followed by a 20-min cool down step. To confirm antibody specificity, slides were incubated with RNaseH (M0297, New England Biolabs)

and RNaseA (EN0531, Thermo Fisher Scientific) enzymes for 1 h and 24 h at 37°C, respectively, as previously described³². Slides were then rinsed in 1 × Tris buffered saline (TBS) three times. Following endogenous peroxidase blocking, the slides were incubated with S9.6 (1:20,000) for 2 h at room temperature in a moist humidity chamber. Anti-mouse Powervision-HRP Conjugated Polymer from Leica (Cat #PV6114) for 30 min was used for detection. Slides were then developed with DAB for 5 min, rinsed with TBS and counterstained with haematoxylin, dehydrated, cleared and mounted with a synthetic mounting medium. Images were taken on an Olympus sc-100 at 60× magnification or a Motic Digital Slide Scanning System at 40× magnification.

Repair assays. The DR-GFP reporter assay was carried out as previously described²². U2OS cells with stably integrated DR-GFP construct and the endonuclease IScel expression vector were obtained from M. Jasin (Memorial Sloan Kettering Cancer Center) and J. Stark (City of Hope Cancer Center). In brief, cells were seeded into 24-well plates and transfected with siRNAs or expression vectors or combinations thereof. On the next day, cells were transfected with IScel expression vector. After 72 h, cells were collected and GFP-positive cells were evaluated by flow cytometry on a BD FACSCanto flow cytometer. Appropriate controls were used and all experiments were performed with transfection triplicates and repeated for independent validation.

BRCA1 ChIP-qPCR at DR-GFP sites. This assay was carried out as previously described³³. In brief, DR-GFP U2OS cells were transfected with EWS-FLI1 or the empty vector control. Twenty-four hours later, a DSB was induced by transfection with IScel expression vector. Cells were fixed with 1% formaldehyde 16 h later and then washed with ice-cold 0.5% BSA/PBS. Fixed cells were resuspended in lysis buffer (1% SDS, 10 mM EDTA, 0.5 mM EGTA, 50 mM Tris-HCl pH 7.5, 0.2% Triton X-100, protease and phosphatase inhibitor cocktail), and sonicated to achieve a desired median fragment length of 200 bases. Lysates were then diluted in immunoprecipitation buffer (1% Triton X-100, 2 mM EDTA, 150 mM NaCl, 20 mM Tris-HCl pH 7.5, protease and phosphatase inhibitor cocktail). Five per cent of the solution was reserved as input. The remaining pre-cleared lysates were used for overnight immunoprecipitation with either BRCA1 antibody (A300-000, Bethyl Labs) or IgG control (ab37415, Abcam) and pre-washed beads. On the next day, beads were washed with SDS-free RIPA/LiCl buffer (50 mM HEPES pH 7.5, 1 mM EDTA, 0.7% sodium deoxycholate, 1% NP-40, 0.5 M LiCl, protease and phosphatase inhibitor cocktail) followed by washes in 1 × TE buffer. Immunoprecipitated beads were resuspended in elution buffer (1% SDS, 0.1 M sodium bicarbonate) and incubated at 65°C to reverse the crosslinks. Eluates were further subjected to proteinase K treatment for 2 h at 37°C and purified using Qiagen. Real-time quantitative PCR (qPCR) was conducted according to the protocol below along with the primers described previously^{34,35}.

Quantitative PCR. qPCR was performed on DNA samples obtained either after chromatin immunoprecipitation or DNA-RNA immunoprecipitation using SYBR qPCR Mix (Applied Biosystems) according to the manufacturer's protocol. All reactions were carried out in technical triplicates. The following primers were used for qPCR performed with technical triplicates: *APOE*: FP: 5'-CCGGTGAGAAGCGCAGTCCG-3'; RP: 5'-CCCAAGCCCGACCCCG AGTA-3'; *PARP8*: FP: 5'-GGGTGTCTCTTAGGCAGAAC-3'; RP: 5'-ATG GAAACCTGTTTGGCTTG-3'; *FEN1*: FP: 5'-CCTCTCGCCCTTAGAAATCG -3'; RP: 5'-TAGACGCTCCTGGAACCTC-3'.

RNAi screens. The RNAi screens with etoposide, bleomycin and MMS were performed as described previously³⁶. Kc167 cells (*Drosophila melanogaster*; 1.2×10^4) were seeded into 384-well plates with Schneider medium and grown at 22°C in a humidified chamber. Each well of a 384-well plate contained 0.25 µg double-stranded RNA (dsRNA) with 22,915 dsRNA representing the whole library. The top 5% 'survival' hits for each damaging agent were calculated (Supplementary Table 2) as previously described³⁶. Detailed analysis of MMS hits has previously been published³⁷.

RNA-seq and gene expression analysis. IMR90, U2OS, TC32, EWS502, CHLA258 and CHLA10 cell lines were grown in 10-cm dishes to 90% confluence. Samples were collected after 0, 6, 12, 18 or 24 h of etoposide exposure (equitoxic doses leading to 65% viability after 72 h) and RNA extracted using Qiagen RNeasy kit. The quality of RNA samples was analysed using an Agilent 2100 BioAnalyzer. Sequencing libraries were prepared from total RNAs according to Illumina's RNA sample preparation protocol. Samples were barcoded, and pooled and sequenced with a HiSeq 2000 system with the 50 bp paired-end protocol, and with targeted read counts around 30 million reads. Only U2OS samples were sequenced with the 50 bp single-end protocol. We used TopHat2 aligner to map paired reads to the UCSC hg19 genome build. To quantify gene expression, we used HTSeq to obtain raw read counts per gene and then converted to RPKM (read per kilobase of gene length per million reads of the library) according to gene length and total mapped read count per sample. Log₂-transformed RPKM measurement was used as gene expression level. Differential expression analysis and functional annotation

classification were conducted using Gene Set Enrichment Analysis³⁸ (Broad Institute) and DAVID³⁹.

Chromatin immunoprecipitation. Cells were grown to ~80% confluence and treated with etoposide (LD65) for six hours. Chromatin was collected after fixing with 1% formaldehyde and sheared to an average length of 200–1,500 bp using a Branson sonicator. Four hundred micrograms of sheared chromatin was added to 50 µl protein G beads that were pre-incubated with 10 µg antibody: RNAPII (GAH-111, Qiagen), BRCA1 (sc-646, Santa Cruz) and IgG (GAH-111, Qiagen or ab37415, Abcam), for overnight incubation. ChIP DNA was eluted by incubating beads in elution buffer (50 mM Tris pH 8.0, 10 mM EDTA and 1% SDS) overnight at 65°C, followed by sequential treatments with RNaseA and proteinase K. The final ChIP DNA was purified by phenol/chloroform extraction and ethanol precipitation. These samples were further sheared using a Covaris sonicator to an average length of 350 bp. Library construction and purification was done following the manufacturer's protocol (MicroPlex Library Preparation kit, Diagenode and AgenCourt Ampure XP, Beckman Coulter). Two control and two Ewing sarcoma cell lines were used for sequencing and repeated for validation.

DNA-RNA immunoprecipitation. DRIP was performed according to a previously published protocol¹⁷. In brief, DNA from a nearly confluent 10-cm dish was obtained using proteinase K followed by phenol/chloroform extraction and ethanol precipitation. DNA was subjected to overnight digestion using a cocktail of restriction enzymes (HindIII, EcoRI, BsrGI, XbaI and SspI, NEB). After being cleaned up with phenol/chloroform/ethanol, 4 µg digested DNA with or without RNaseH pre-treatment was used as input for immunoprecipitation using S9.6 antibody (Kerafast). The DNA-antibody complex was incubated for 16 h and allowed to bind protein A/G beads for a further 4 h. Bound DNA fragments were recovered in the elution buffer by incubating with proteinase K at 55°C for 45 min. Recovered DNA was cleaned using phenol/chloroform/ethanol and resuspended in 10 mM Tris-HCl pH 8.0. Each immunoprecipitation was run in triplicate and samples were pooled for sequencing after sonication.

ChIP and DRIP sequencing and peak identification. Sonicated and size-selected DNAs (immunoprecipitated DNA that was untreated or treated with etoposide or RNaseH, and input DNA) were processed according to the Illumina Genome DNA library preparation protocol, and sequenced with a HiSeq 2000 or a HiSeq 3000 system with 50 bp single-read sequencing protocol. On average, 30–40 million reads were generated for each DNA sample, and then aligned to the UCSC hg19 genome build using BWA. Peak calling was performed using the MACS2⁴⁰ algorithm. Similar to ref. 17, we used a peak calling parameter with fivefold up to 30-fold enrichment over corresponding input DNA as control (MACS2 parameters: -g 2.7e9 -q 0.05 -B -m 5 30).

Determination of consensus ChIP and DRIP regions. The methodology for analysis has previously been described in detail⁴¹. In brief, DRIP regions were first stacked according to their genomic position (within each chromosome), and then regions that were present in at least three DRIP samples (under any condition) were selected. If the next adjacent region was less than 200 bp away, we treated them as being contiguous. Regions smaller than 200 bp were eliminated. With each seed region, we extended it to the longest run within the stack, followed by another merging step between adjacent regions that were less than 200 bp apart. A total of 33,121 DRIP regions were detected. Small DRIP regions were determined by removing samples with strong DRIP peaks (EWS502, CHLA10, and TC32), and then performing the consensus peak detection algorithm again.

ChIP and DRIP region coverage. We quantified DRIP regions by coverage and sequence read count. DRIP regions with greater coverage and higher enrichment (large sequence read count), were indicative of a true DRIP peak. After obtaining the consensus regions, coverage of each sample was defined as:

Coverage =

$$\frac{\sum_{\text{all peaks}} \text{Overlap of detected DRIP peaks with consensus regions (bp)}}{\sum_{\text{all consensus peaks}} \text{Consensus regions (bp)}}$$

In other words, the coverage of a given sample is the percentage of consensus peaks that are covered by the original DRIP peaks obtained by the MACS algorithm.

Normalized read counts. To quantify the read counts within the DRIP regions, we first counted short sequence reads within 100-bp tiling bins equally across the entire genome using the BedTools/CoverageBed command. The normalized read count of a given DRIP region was quantified as:

$$\text{Readcount} = \frac{(1/N) \sum_{\text{all bins overlap with consensus DRIP regions}} \text{reads in each bin} \times 50/100}{\text{Total number of mapped reads}/10,000,000}$$

in which N is the total number of DRIP peaks. Read count per DRIP region was normalized to 10 million reads per library. By considering read length (50 bp) and bin size (100 bp), the read count unit essentially defines the normalized depth of coverage per 10 million reads per library. Therefore, if we use consensus DRIP regions for all samples, the read count of a given DRIP region becomes comparable between samples.

Gene sets derived from EWS ChIP-seq or other profiling techniques. EWS-FLI1 targeted genes determined by ChIP-seq were extracted from ref. 18. One thousand, seven hundred and eight-five EWS-FLI1 binding sites were reported, which we mapped to 1,314 unique genes.

DRIP-seq and ChIP-seq heat maps. For any given gene set, a heat map of DRIP data, ChIP data and gene expression data was generated for correlation examination. The gene set was provided in a BED format that contains at least five columns: 1) chromosome; 2) genomic start position; 3) genomic end position; 4) unique ID; 5) gene symbol. All genomic data were derived from the UCSC hg19 genome build. If defined by TSS, columns 2 and 3 were ± 100 bp around the TSS defined by the farthest TSSs, when multiple transcripts are defined in the UCSC hg19 'refFlat' table. If the genes were derived from ChIP-seq data, then the genomic positions were simply the DNA binding sites from a given pull-down target. Heat map selects 50 bins (5,000 bp) to the left and right (x -axis) for a given gene set (y -axis), then plots the read count value from white (read count 0) to dark red. Gene set (y -axis) can be sorted by the peak value defined in the following sections. **DRIP-seq.** Raw sequence read counts around $\pm 5,000$ bp (50 bins to each side, with bins of 100 bp) were extracted around the centre of each site defined in the gene set. Each gene was also represented by a mean read count over 100 bins. If a specific order for DRIP data was requested, it was sorted again by read count. To generate a colour map, we used a colour scheme as follows: white, no reads; magenta, with detectable read counts; black/dark magenta, height peaks. DRIP-seq data were normalized to 20 million reads per library equivalent.

ChIP-seq. Similar to DRIP-seq data, raw read counts $\pm 5,000$ bp were extracted around the centre of each site defined in the gene set. Given that BRCA1 binding sites are mostly narrow peaks, we calculated the mean peak height over windows of seven consecutive bins (700 bp), and then took the maximum height over all moving seven-bin windows, or:

$$\text{peak height} = \max_{j=3 \text{ to } 47} \left(\frac{1}{7} \sum_{k=-3}^3 \text{RC}_{j+k} \right)$$

in which RC is the read count of a given bin. By doing so, we allow the ChIP-seq peak to be within the region, but not necessary situated at the centre, particularly when TSSs or TTSs were requested. Heat maps were generated using the same colour scheme as for DRIP-seq. The code for rendering the heat maps was written in MATLAB and will be made available upon request.

Gene expression. Gene expression data were extracted using gene symbol matching. Log₂-transformed RPKM was used to represent expression level. The heat map colour scheme is as follows: magenta, higher expression; blue, lower expression.

Kolmogorov-Smirnov test for enrichment. For a given gene list, we sorted ChIP data, gene expression data, DRIP, RNAPII and BRCA1 ChIP data (as shown in Extended Data Fig. 4c) according to ChIP-seq peak height, gene expression level or DRIP-seq mean read count. To demonstrate the concordance of ChIP peaks and DRIP peaks, for example, we hypothesized that if concordance did not exist, then

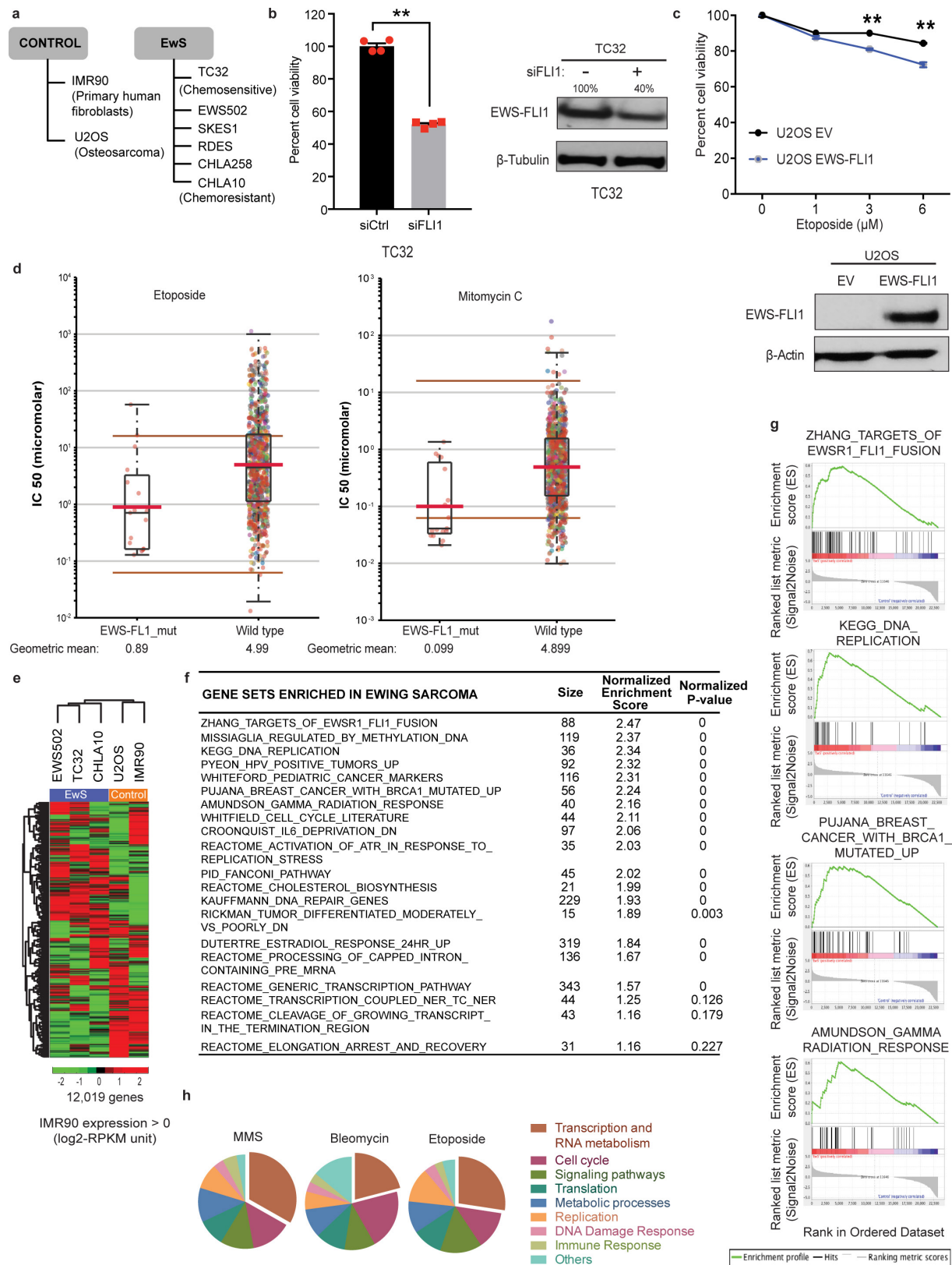
the DRIP peaks would be uniformly distributed within the gene list. Therefore, we first determined the ChIP peak threshold (peak height > 9), DRIP peak threshold (read count $> \mu + \text{MAD} \times 1.89$, in which μ is the median threshold obtained by calculating mean values across a ± 5 -kb region surrounding the peak and MAD is the median absolute deviation for the same) and expression > 5 . Depending on the number of peaks we got from ChIP-seq (for example), we determined the threshold for expression and DRIP-seq such that both will have the same positive number of peaks or genes, and we then examined their location within gene list and determined whether they are uniformly distributed using the Kolmogorov-Smirnov test. The code for performing the statistical test was written in MATLAB and will be made available upon request.

Statistical analysis. P values for analysing cell viability, R-loop intensity, homologous recombination repair and enzyme assays were computed using either Student's t -test (two-tailed) or one-way ANOVA with Bonferroni correction for cell line differences at each drug dose in GraphPad Prism software. Where applicable, two-way ANOVA was employed. When performing multiple comparisons, an FDR of 1% was used as cutoff as evaluated using the Benjamini-Hochberg method. $P < 0.05$ or 0.005 was considered significant (marked as *, # or **, ## respectively). No statistical methods were used to predetermine sample size.

Code availability. All custom MATLAB codes are available from the corresponding author upon request.

Data availability. RNA-seq and DRIP-seq datasets have been deposited at the Gene Expression Omnibus (GEO) with accession code GSE68847. For gel source data, see Supplementary Fig. 1. All other data and are available from the corresponding author upon request.

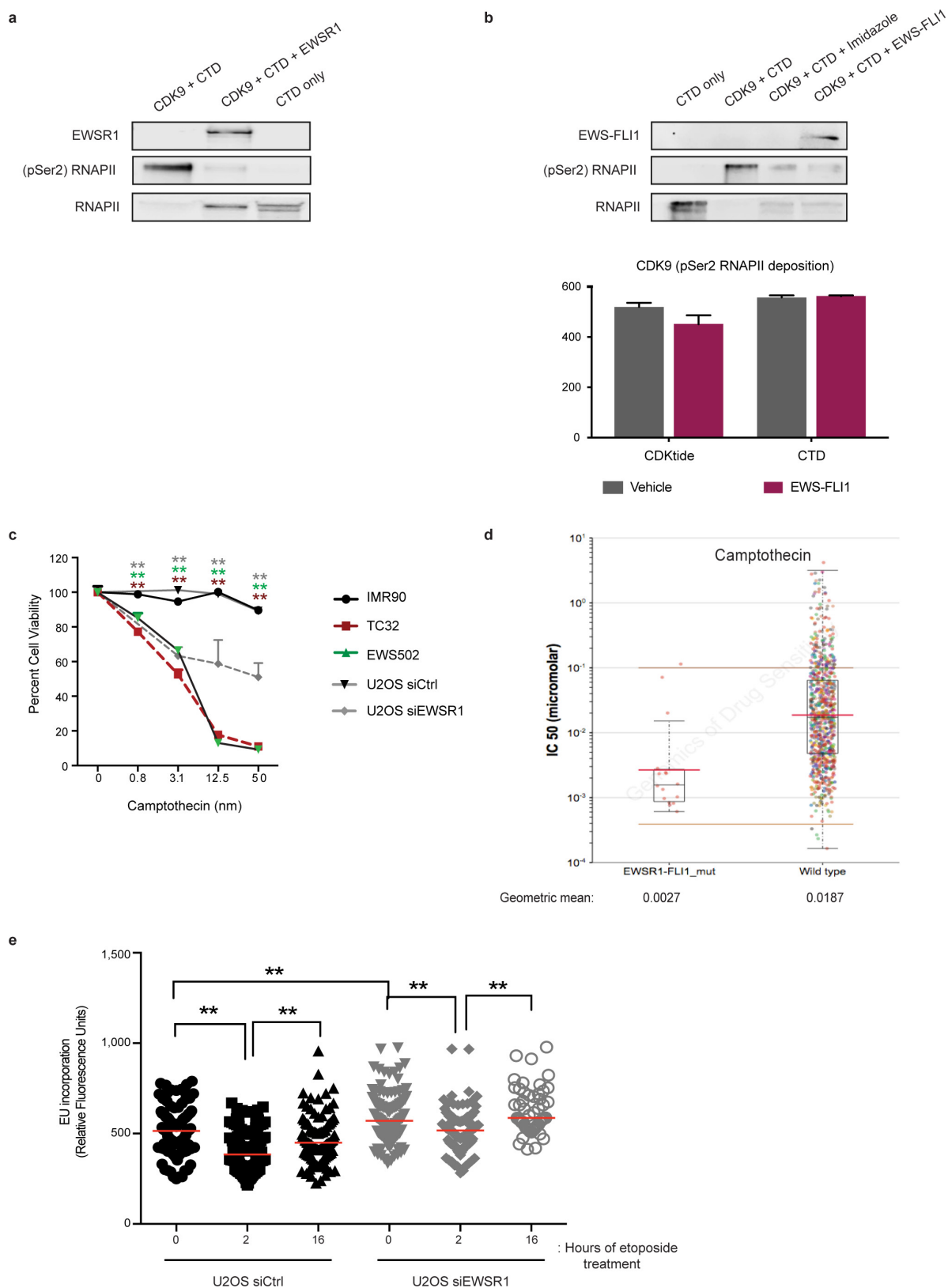
- Suzuki, Y. *et al.* An upstream open reading frame and the context of the two AUG codons affect the abundance of mitochondrial and nuclear RNase H1. *Mol. Cell. Biol.* **30**, 5123–5134 (2010).
- Sanz, L. A. *et al.* Prevalent, dynamic, and conserved R-loop structures associate with specific epigenomic signatures in mammals. *Mol. Cell* **63**, 167–178 (2016).
- Hu, Y. *et al.* PARP1-driven poly-ADP-ribosylation regulates BRCA1 function in homologous recombination-mediated DNA repair. *Cancer Discov.* **4**, 1430–1447 (2014).
- De Luca, P. *et al.* BRCA1 loss induces GADD153-mediated doxorubicin resistance in prostate cancer. *Mol. Cancer Res.* **9**, 1078–1090 (2011).
- Gorski, J. J. *et al.* Profiling of the BRCA1 transcriptome through microarray and ChIP-chip analysis. *Nucleic Acids Res.* **39**, 9536–9548 (2011).
- Wiles, A. M., Ravi, D., Bhavani, S. & Bishop, A. J. An analysis of normalization methods for Drosophila RNAi genomic screens and development of a robust validation scheme. *J. Biomol. Screen.* **13**, 777–784 (2008).
- Ravi, D. *et al.* A network of conserved damage survival pathways revealed by a genomic RNAi screen. *PLoS Genet.* **5**, e1000527 (2009).
- Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
- Huang, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protocols* **4**, 44–57 (2009).
- Zhang, Y. *et al.* Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137 (2008).
- Liu, R. *et al.* Computation tools for genome-wide R-loops identification and characterisation. *Int. J. Comput. Biol. Drug Des.* **10**, 123–136 (2017).



Extended Data Figure 1 | See next page for caption.

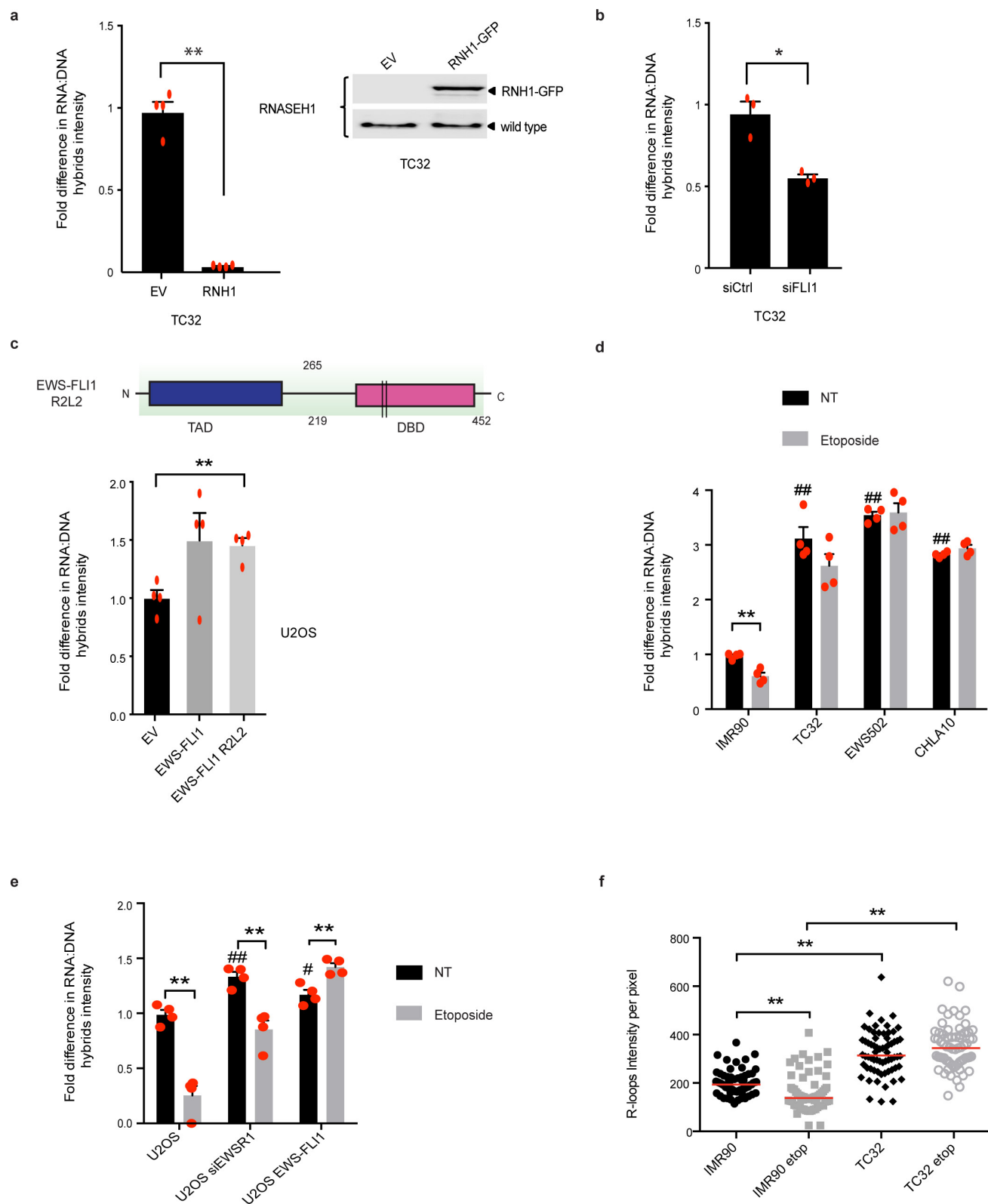
Extended Data Figure 1 | Characterizing Ewing sarcoma chemosensitivity. **a**, Cell lines used in the study. **b**, Level of cell death caused by EWS-FLI1 knockdown alone in TC32 cells. Immunoblot shows extent of knockdown; $n = 4$ transfection replicates. **c**, Cell viability of U2OS cells transfected either with empty vector (EV) or EWS-FLI1 for 24 h before etoposide exposure for a further 48 h. Immunoblot shows transfection efficiency; $n = 3$ transfection replicates. **d**, IC_{50} levels of etoposide or mitomycin in EWS-FLI1 mutant ($n = 16$) versus pan-cancer ($n = 143$) dataset. Brown lines, range of screening concentrations of the drug. Red lines, geometric mean of drug concentration. **e**, Heat map of basal gene expression profile in control and Ewing sarcoma cell lines

after hierarchical clustering. **f**, **g**, Top enriched pathways from gene set enrichment analysis of the differences between Ewing sarcoma and IMR90 cells are listed (**f**) and relevant signature plots are illustrated (**g**). We found differential upregulation of replication stress, BRCA1-mutation driven network and altered transcription regulation pathways in Ewing sarcoma. **h**, Cross-screen pathway comparison of top survival hits from RNAi screens in *Drosophila* Kc167 cells exposed to MMS, bleomycin or etoposide. Nearly a third of the top 5% hits in each screen were genes involved in transcription and RNA metabolism, highlighting the importance of this pathway in DNA damage survival. Mean \pm s.e.m., $*P < 0.05$, $**P < 0.005$, two-tailed t -test.



Extended Data Figure 2 | Aberrant transcription regulation in Ewing sarcoma. **a**, Immunoblot depicting the phosphorylation of RNAPII CTD fragment used as the substrate in Fig. 1d. Recombinant EWSR1 and hypophosphorylated RNAPII are also displayed. **b**, Level of inhibition of CDK9 activity by buffer (vehicle) or recombinant EWS-FLI1 protein on the two RNAPII CTD substrates. The immunoblot on top is confirmation of kinase activity measured by the assay. **c**, Cytotoxicity profile in response to camptothecin in control, Ewing sarcoma and EWSR1-depleted cells.

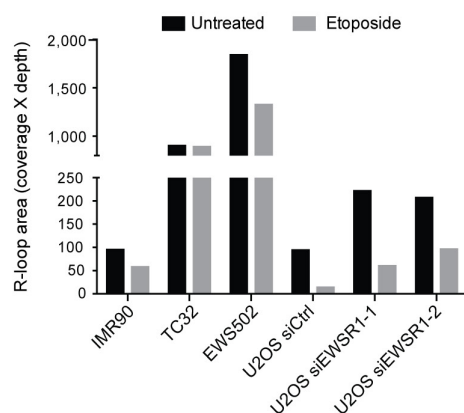
$n = 4$ technical replicates, one-way ANOVA against IMR90 cells. **d**, IC₅₀ levels of camptothecin in EWS-FLI1 mutant ($n = 15$) versus pan-cancer ($n = 132$) dataset. Brown lines, range of screening concentrations of the drug. Red lines, geometric mean of drug concentration. **e**, Transcription restart assay measured in U2OS cells transfected with either scrambled or EWSR1 siRNA. $n = 4$ transfection replicates, two-way ANOVA. Mean \pm s.e.m., * $P < 0.05$, ** $P < 0.005$.



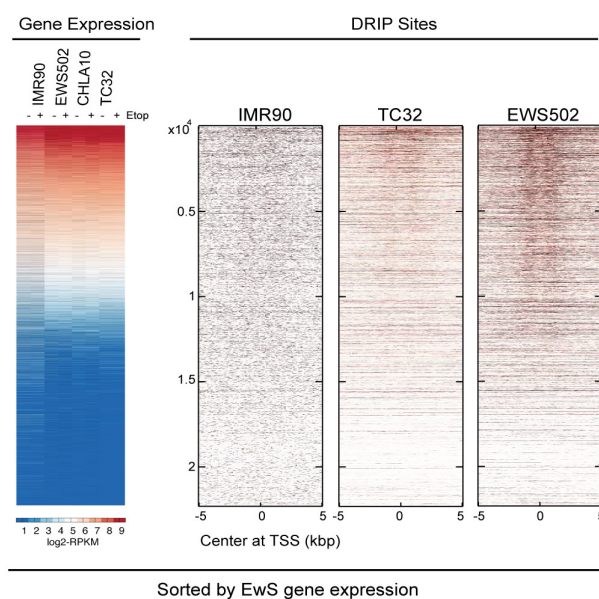
Extended Data Figure 3 | R-loops in Ewing sarcoma. **a**, Quantification of RNA–DNA hybrids in TC32 cells transfected with either empty vector (EV) or RNaseH1 (RNH1). The immunoblot to the right indicates RNaseH1 transfection efficiency; $n = 4$ transfection replicates. **b**, RNA–DNA hybrid levels in TC32 cells with scrambled (siCtrl) or EWS–FLI1 (siFLI1) knockdown; $n = 3$ transfection replicates. **c**, Schematic of the EWS–FLI1 R2L2 construct. Arginine residues 383 and 386 (black bars) in EWS–FLI1 are converted to leucine to render the fusion oncogene deficient in DNA binding. Below is a quantification of RNA–DNA hybrids in U2OS cells expressing empty vector, EWS–FLI1 or EWS–FLI1 R2L2;

$n = 4$ transfection replicates. **d**, **e**, Fold change in RNA–DNA hybrids after damage (etoposide, 6 h) in IMR90 versus Ewing sarcoma cells (**d**) or U2OS cells with either EWSR1 depletion or EWS–FLI1 expression (**e**). NT, no treatment; $n = 4$ technical or transfection replicates. **f**, Quantification of nucleoplasmic RNA–DNA hybrids in the immunofluorescence images ($n = 80$ nuclei) demonstrated a clear increase in overall R-loop intensity in Ewing sarcoma cells. Nucleolin signal was used to subtract nucleolar R-loops. One-way ANOVA. Mean \pm s.e.m., *, $P < 0.05$, ##, $P < 0.005$. # indicates significance of Ewing sarcoma relative to untreated IMR90 cells or transfections relative to U2OS cells.

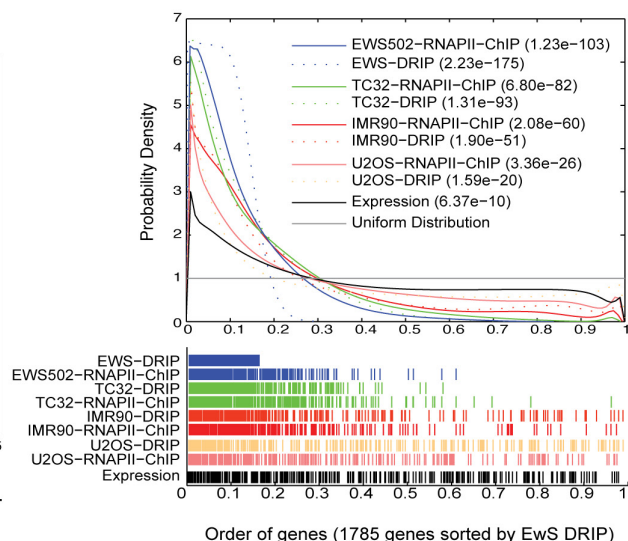
a



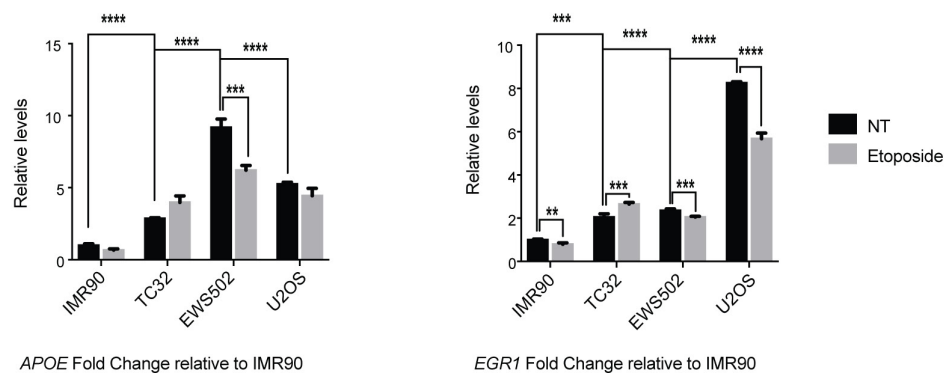
b



c

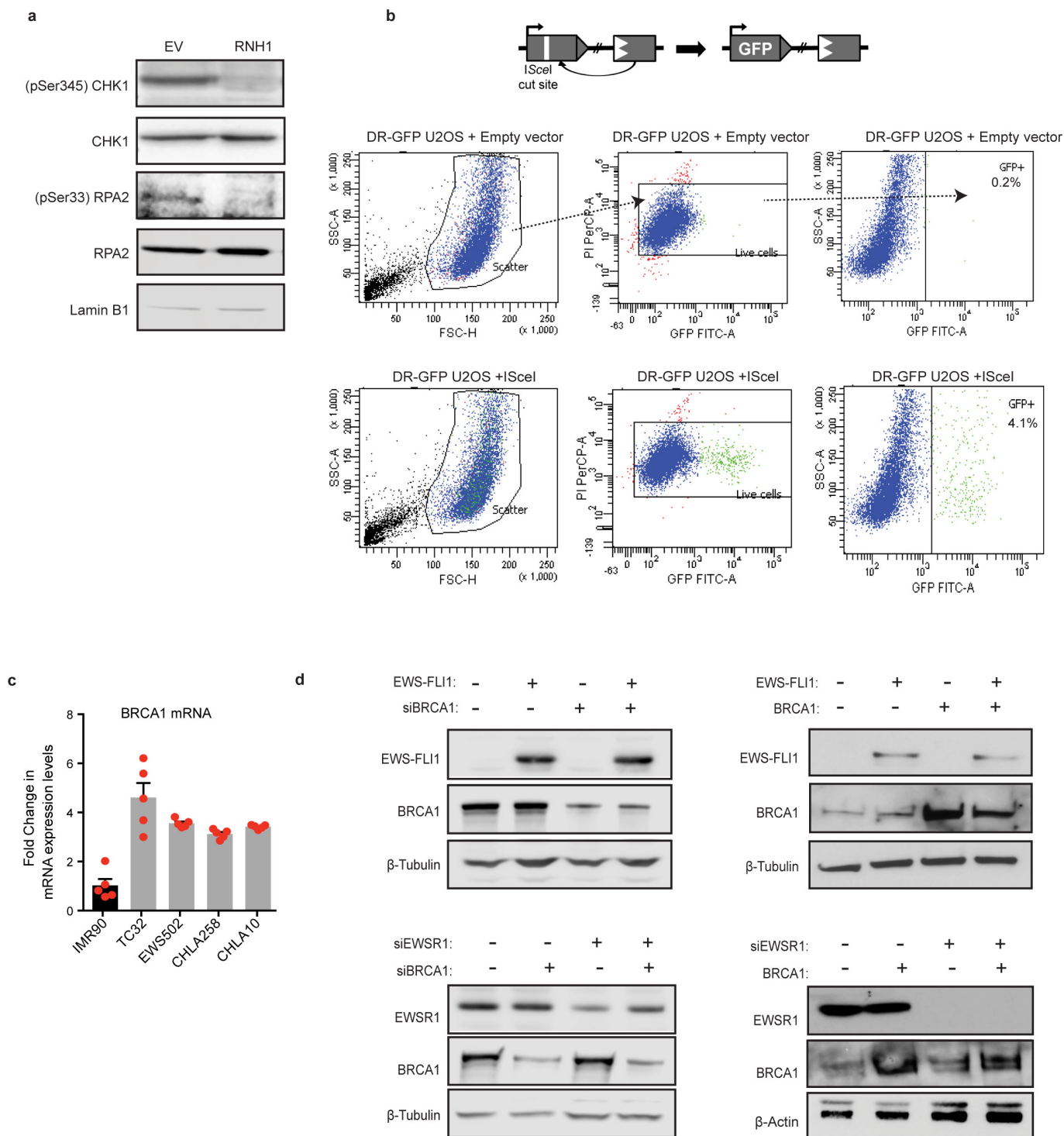


d



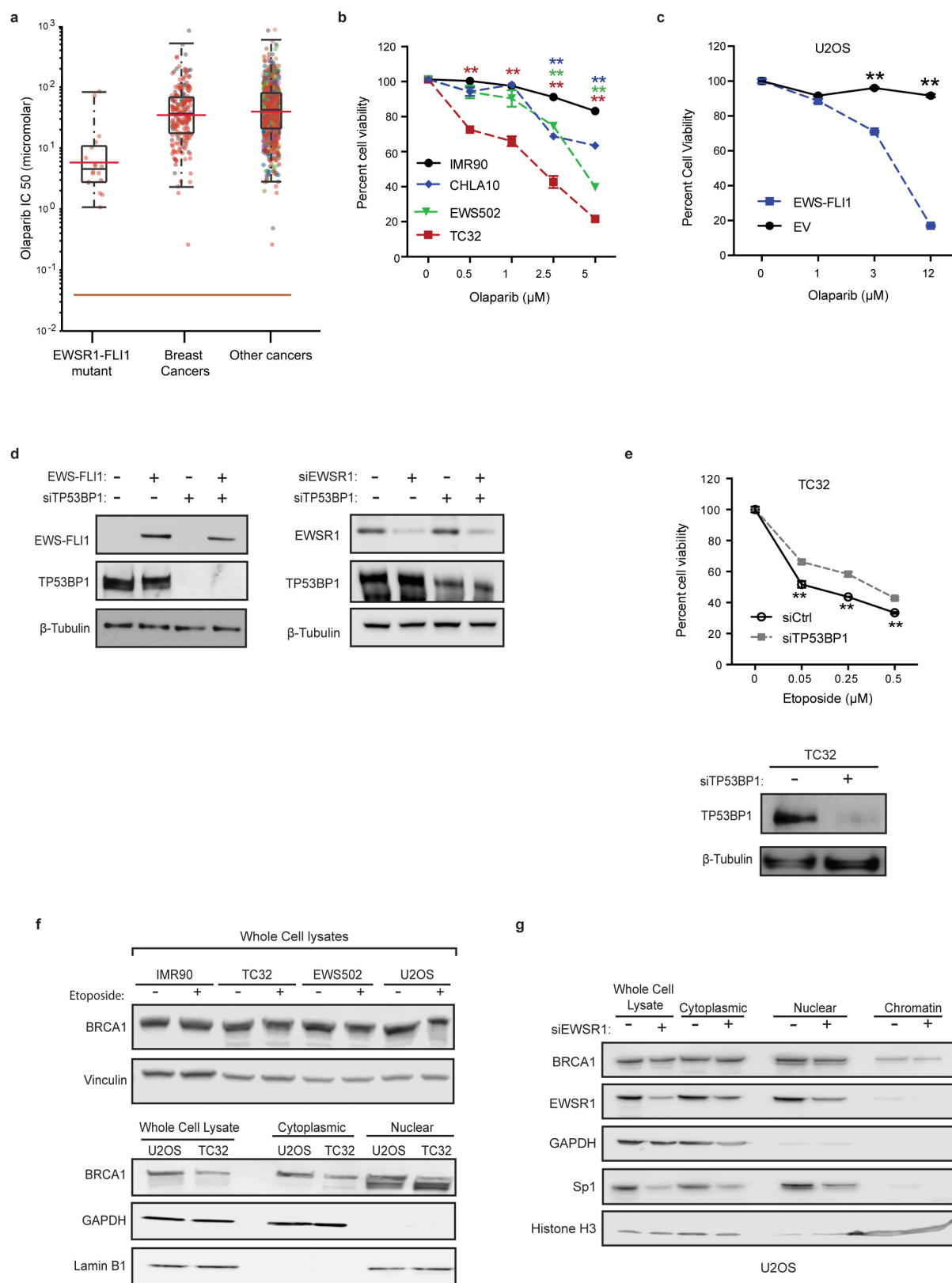
Extended Data Figure 4 | DRIP-seq data validation. **a**, Quantification of DRIP (coverage of DRIP region multiplied by reads in that region) across all samples. y-axis is graphed in logarithmic scale. **b**, Representative whole-genome heat maps centred around TSS ordered by average expression of Ewing sarcoma cells. **c**, Probability density graph plotted with a Gaussian smoothing kernel of the distribution of DRIP peaks and EWS-FLI1 ChIP peaks at EWS-FLI1 bound genes relative to uniform distribution. $n = 281$

genes (top 16%). Inset, P values depicting significance of enrichment for each sample. **d**, Fold enrichment of qPCR product from ChIP experiments done with RNAPII antibody in control and Ewing sarcoma cell lines. The primers target well-known R-loop regions within *APOE* and *EGR1* genes. Mean \pm s.e.m., $n = 3$ technical replicates, *** $P < 0.0005$, **** $P < 0.00005$. One-way ANOVA across cell lines compared to IMR90 cells and two-tailed t -test within cell lines.



Extended Data Figure 5 | R-loop-dependent replication stress and recombination defect in Ewing sarcoma. **a**, Representative immunoblots evaluating decrease in ATR kinase pathway activation upon overexpression of RNaseH1 in TC32 cells. **b**, Schematic of the DR-GFP construct integrated into U2OS cells. Below are representative scatter plots

of the gating scheme used to determine percentage of GFP-positive cells after inducing a DSB via ISceI vector compared to empty vector. **c**, RNA-seq data of BRCA1 transcript levels in Ewing sarcoma cell lines compared to IMR90 cells. **d**, Immunoblots demonstrating transfection efficiency of indicated siRNA and expression constructs used in Fig. 3f, g.

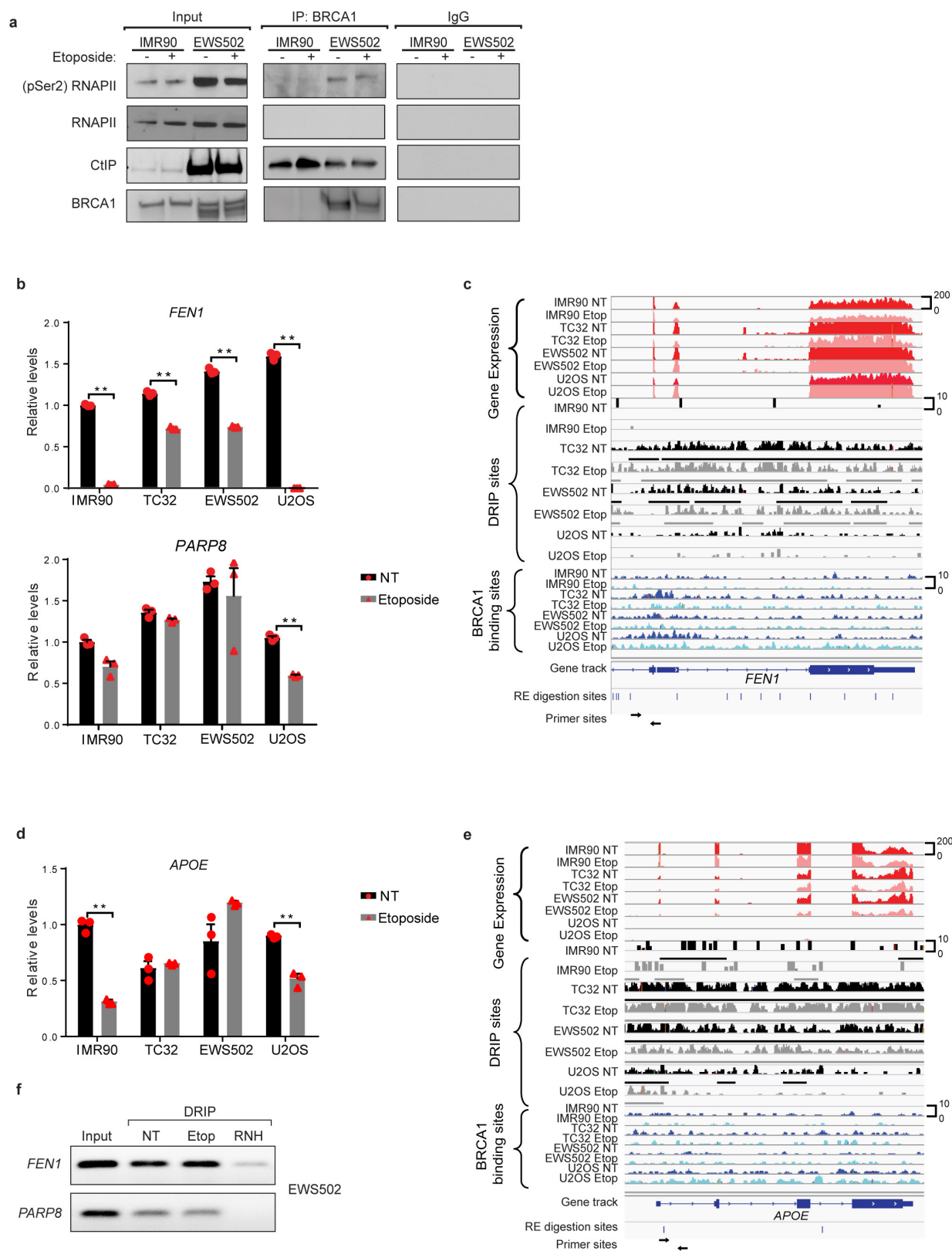


Extended Data Figure 6 | See next page for caption.

Extended Data Figure 6 | Similarity of Ewing sarcoma to BRCA-deficient tumours.

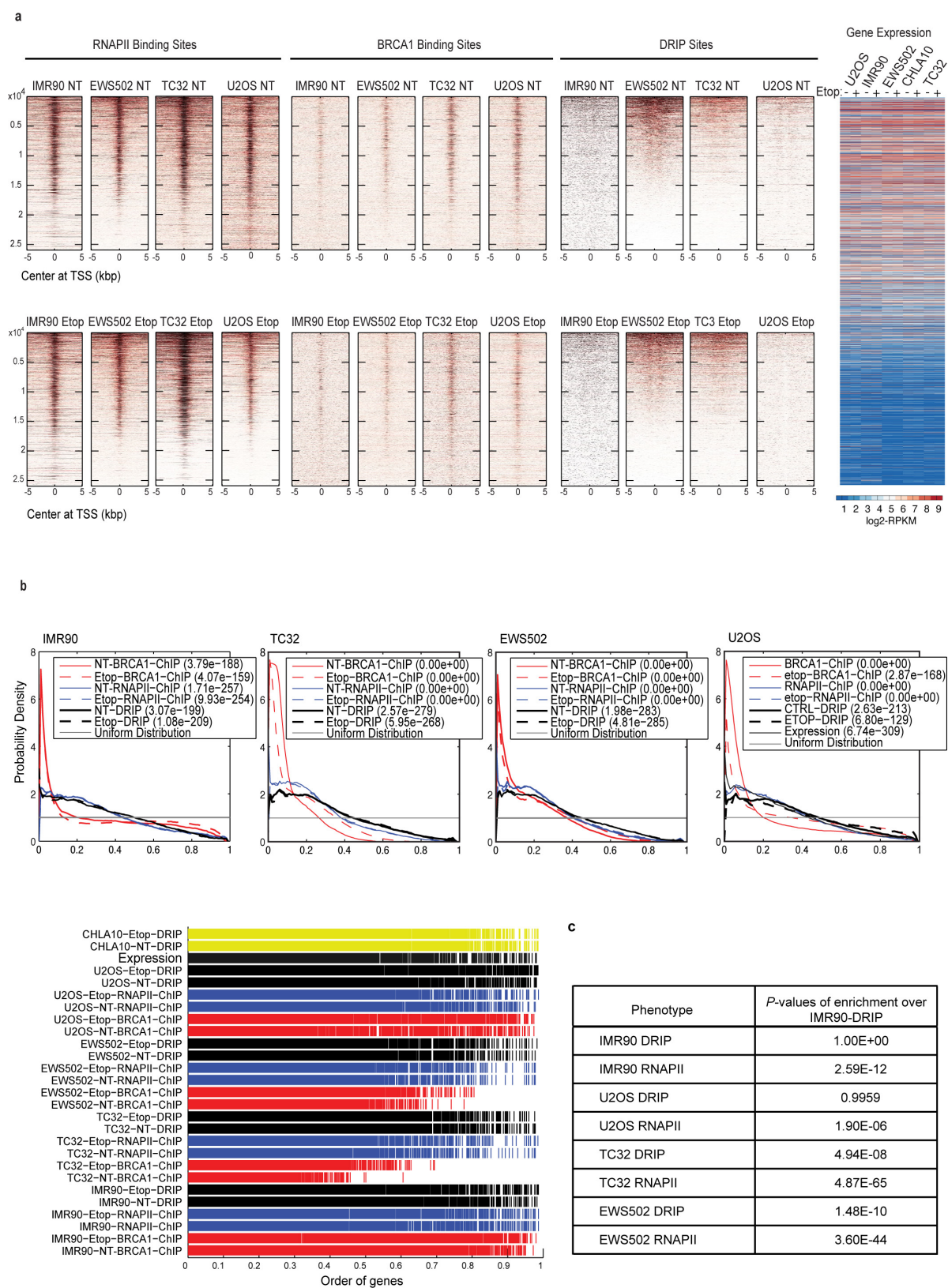
a, IC₅₀ levels of olaparib in EWS-FLI1 mutant cells ($n = 17$) versus breast cancers ($n = 13$) or pan-cancer ($n = 147$) dataset. **b**, Cell viability of IMR90 and Ewing sarcoma cells with increasing doses of olaparib. Mean \pm s.d., $n = 3$ technical replicates, one-way ANOVA compared to IMR90 cells. **c**, Cell viability plot demonstrating the role of EWS-FLI1 in mediating exquisite sensitivity to olaparib in U2OS cells transfected with either the oncogene or empty vector; $n = 3$ transfection replicates. **d**, Immunoblots depicting transfection efficiency of indicated siRNA and expression constructs used in Fig. 3h. **e**, TP53BP1 knockdown improved Ewing sarcoma (TC32 cell) survival in response to damage. Immunoblots depict level of TP53BP1 knockdown. $n = 4$ transfection

replicates. **f**, Representative immunoblots showing equivalent levels of BRCA1 in whole cell lysates (upper panel) from control and Ewing sarcoma cells with and without etoposide treatment (2 h). The lower panel shows BRCA1 redistribution in subcellular fractions of U2OS or TC32 cells. GAPDH and lamin B1 were used as loading controls for the cytoplasmic and nuclear fractions, respectively. **g**, Immunoblots of whole cell lysates and subcellular fractions from U2OS cells with and without EWSR1 depletion. Data indicated no change in BRCA1 levels with EWSR1 knockdown. Loading controls include: GAPDH for cytoplasm, Sp1 for nuclei and histone H3 for chromatin. Mean \pm s.e.m., $**P < 0.005$, two-tailed t -test at each dose.



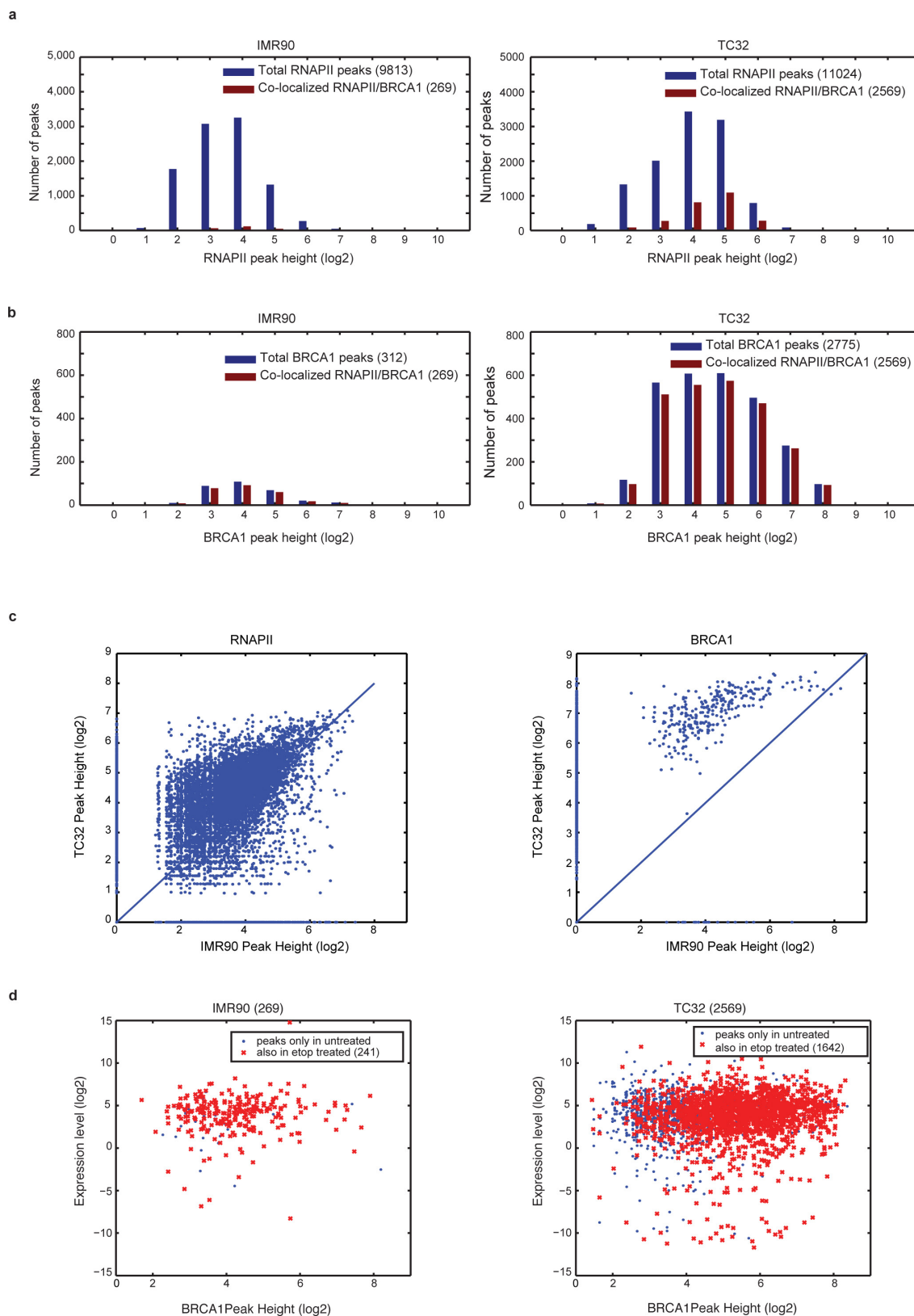
Extended Data Figure 7 | Association of BRCA1 with the transcription complex in Ewing sarcoma. **a**, Co-immunoprecipitation: immunoblots of IMR90 and EWS502 nuclear lysates with and without exposure to etoposide (2 h). The left panel indicates 10% of the input used for immunoprecipitation. BRCA1 antibody was used for immunoprecipitation in the middle panel and the rightmost panel indicates specificity of interaction against IgG pulldown. **b**, Real-time qPCR analysis of BRCA ChIP samples from control and Ewing sarcoma cell lines with and without etoposide treatment, using primers within the *FEN1* and *PARP8* genes. **c**, Representative sequencing track image of gene expression (red tracks),

R-loop sites (black and grey tracks) and BRCA1 binding sites (blue tracks) across the *FEN1* gene demonstrating the enrichment of R-loops and BRCA1 in the region amplified by the primers in **b**. **d**, qPCR analysis as in **b** with primers targeting a well-known R-loop region within the *APOE* gene. **e**, Representative sequencing track image as in **c** across the *APOE* gene. **f**, Agarose gel blots evaluating amplicons generated using EWS502 DRIPs with primers against *FEN1* and *PARP8*. NT, no treatment; Etop, etoposide-treated (6 h); RNH, RNaseH-treated samples. Mean \pm s.e.m., $n = 3$ technical replicates, $**P < 0.005$, two-tailed *t*-test.



Extended Data Figure 8 | Genome-wide heat maps. **a**, Heat maps representing genome-wide localization of RNAPII, BRCA1 and R-loop sites centred on the TSS. The data were sorted by DRIP sites. The upper panel represents untreated (NT) samples and the lower panel represents etoposide (Etop, 6 h) treated samples. There was a clear decrease in BRCA1 and R-loop signal upon damage in the control cell lines, unlike in Ewing sarcoma. **b**, KS plots to demonstrate empirical distribution of the top

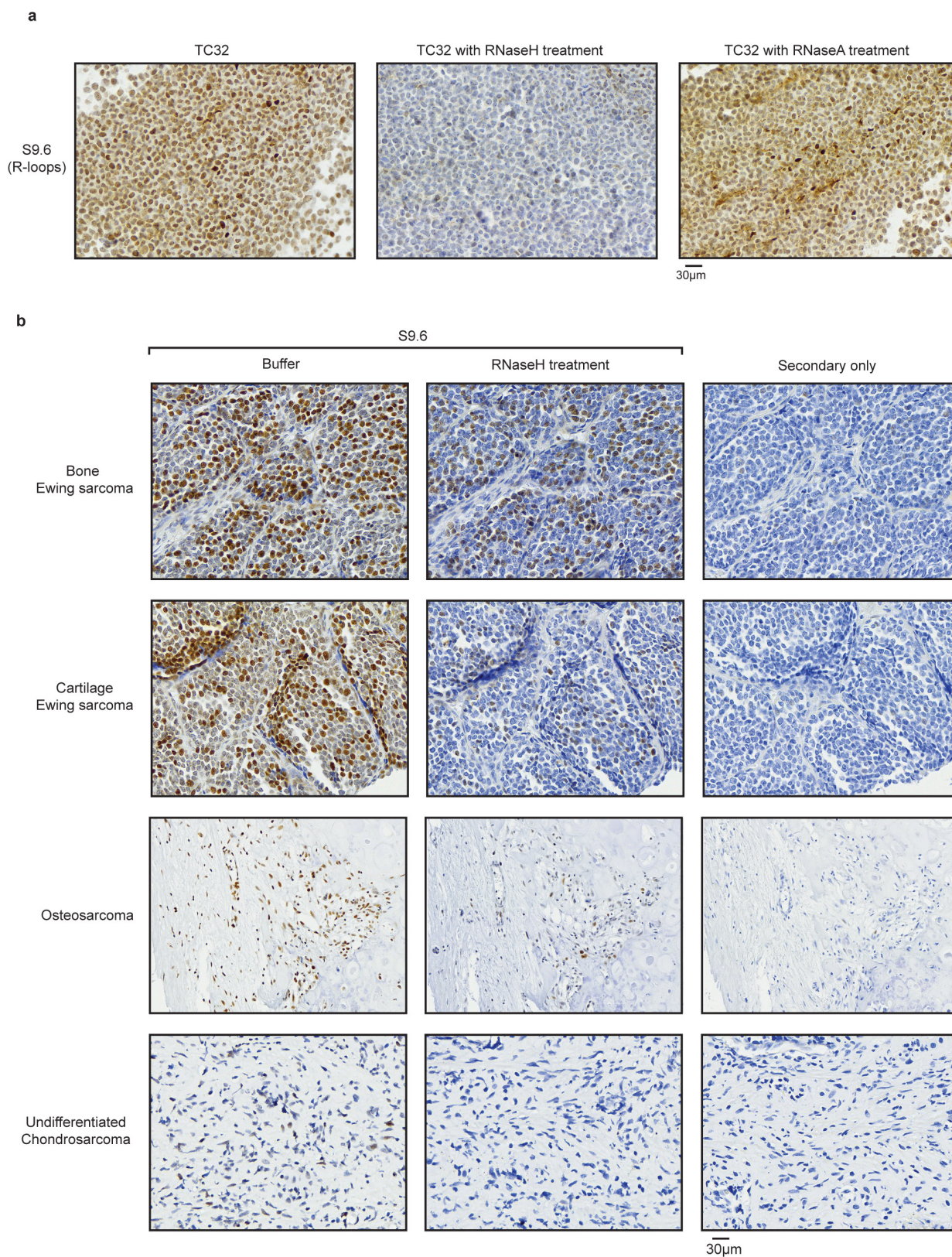
13.8% of DRIP and ChIP peaks and higher expression relative to uniform distribution. Data are sorted by BRCA1 ChIP, $n = 3,066$ genes. **c**, P values of statistical comparisons between RNAPII ChIP and R-loop probability distributions for all cell lines against IMR90 DRIP data centred on the TSS. The top 27% of DRIP-seq peaks corresponding to 6,127 genes were used for the analysis and data were sorted by BRCA1 binding sites.



Extended Data Figure 9 | See next page for caption.

Extended Data Figure 9 | Correlation between BRCA1 and RNAPII binding. **a**, Distribution of RNAPII abundance across the genome for IMR90 and TC32 cells. The bars depict the number of RNAPII bound sites as a function of the number of peaks (y -axis) and relative abundance (peak height, log-transformed) within these peaks (x -axis). The blue bars indicate the total number of peaks determined from RNAPII ChIP-seq and the red bars represent the peaks that co-localize with BRCA1 peaks obtained from BRCA1 ChIP-seq. The data indicated a similar number of RNAPII peaks for TC32 (11,024) compared to IMR90 (9,813), but a greater amount of DNA bound at these peaks, implying increased RNAPII binding. Furthermore, a higher proportion of RNAPII-bound loci also co-localized to BRCA1 binding sites (red bars) in TC32 than in IMR90 cells (23% compared to 2.7%) and there was a clear increase in RNAPII abundance at these sites. **b**, Distribution of BRCA1 abundance across the genome for IMR90 and TC32, similar to **a**. The data indicate a significantly higher number of total BRCA1 peaks in TC32 cells as well as

a significantly higher enrichment of BRCA1 within these peaks in TC32 cells compared to IMR90 cells. The data also suggest that the majority of the BRCA1 peaks were co-localized with RNAPII. **c**, Scatter plots represent the correlation of RNAPII (left) and BRCA1 (right) peak heights between TC32 and IMR90 cells. Data were plotted after being normalized to read count and log-transformed to make comparisons. Loci that are unique to each cell line map to the axes whereas common loci are scattered around the diagonal. The data clearly suggest an increase in enrichment of both RNAPII and BRCA1 in TC32 cells compared to IMR90 cells. **d**, Scatter plots represent the relationship between co-localized BRCA1 and RNAPII peaks as a function of BRCA1 peak height (x -axis) and level of expression of the gene associated with these binding sites (y -axis). TC32 cells showed a greater than fivefold increase in the number of BRCA1 peaks that were associated with RNAPII at highly expressed genes. Further, as in **b**, there was a greater abundance of BRCA1 (peak height) at these highly expressed genes in TC32 cells than in IMR90 cells.



Extended Data Figure 10 | Immunohistochemical analysis of tissue sections. a, Representative images depicting R-loop staining by S9.6 antibody on sections derived from fixed TC32 cell pellets. Sections were treated with buffer (left), RNaseH (middle) or RNaseA (right) after antigen retrieval. The slides demonstrate loss of R-loop signal after treatment with RNaseH, as expected. RNaseA treatment, which at higher salt concentrations specifically cleaves single-stranded RNA, did not result in a significant loss of R-loop signal confirming the specificity of

S9.6 antibody in detecting RNA–DNA hybrids. **b,** Representative images from a pan-sarcoma tissue microarray. The left and centre panels were probed with S9.6 antibody with or without RNaseH treatment. The right panel was stained with secondary antibody alone and serves as a non-specific control. Each row represents images from one tumour indicated on the left. Images were scanned at 40 \times (bar at the bottom right denotes resolution).

Extended Data Table 1 | Comparison of Ewing sarcoma to normal cells and BRCA-mutant cancers.**a**

Processes upregulated in IMR90 by etoposide but not in EwS cell lines

Biological Process	P-value
Regulation of cell proliferation	3.895E-10
Regulation of programmed cell death	4.607E-07
Regulation of phosphorylation	2.1199E-07
Positive regulation of gene expression	0.0193

b

Processes downregulated in IMR90 by etoposide but not in EwS cell lines

Biological Process	P-value
Cell cycle	1.51E-35
Mitosis	4.93E-34
Positive regulation of macromolecular metabolic process	9.08E-05
Regulation of gene expression	1.62E-04
Regulation of RNA metabolic processes	2.38E-03

c

Phenotype	BRCA1 mutant cancers	Ewing sarcoma
Sensitivity to DNA damaging agents	✓	✓
Sensitivity to PARP1 inhibitors	✓	✓
Hyperactivation of PARP1	✓	✓
Impaired homologous recombination	✓	✓
Accumulation of R-loops	✓	✓
High expression of EZH2	✓	✓

Significantly enriched pathways obtained from classification based on Gene Ontology annotation of biological processes of genes that were differentially altered between Ewing sarcoma and IMR90 cells. **a**, Differentially upregulated processes in IMR90 but not Ewing sarcoma cells. **b**, Differentially downregulated processes in IMR90 but not Ewing sarcoma cells. **c**, Characteristic hallmarks of BRCA1-mutant breast cancers and whether these phenotypes are similarly observed in Ewing sarcoma.

Epigenetic reprogramming enables the transition from primordial germ cell to gonocyte

Peter W. S. Hill^{1,2}, Harry G. Leitch^{1,2*}, Cristina E. Requena^{1,2*}, Zhiyi Sun^{3*}, Rachel Amouroux^{1,2*}, Monica Roman-Trufero^{1,2}, Malgorzata Borkowska^{1,2}, Jolyon Terragni³, Romualdas Vaisvila³, Sarah Linnett^{1,2}, Hakan Bagci^{1,2†}, Gopuraja Dharmalingham^{1,2}, Vanja Haberle^{1,2†}, Boris Lenhard^{1,2}, Yu Zheng³, Sriharsa Pradhan³ & Petra Hajkova^{1,2}

Gametes are highly specialized cells that can give rise to the next generation through their ability to generate a totipotent zygote. In mice, germ cells are first specified in the developing embryo around embryonic day (E) 6.25 as primordial germ cells (PGCs)¹. Following subsequent migration into the developing gonad, PGCs undergo a wave of extensive epigenetic reprogramming around E10.5–E11.5^{2–11}, including genome-wide loss of 5-methylcytosine^{2–5,7–11}. The underlying molecular mechanisms of this process have remained unclear, leading to our inability to recapitulate this step of germline development *in vitro*^{12–14}. Here we show, using an integrative approach, that this complex reprogramming process involves coordinated interplay among promoter sequence characteristics, DNA (de)methylation, the polycomb (PRC1) complex and both DNA demethylation-dependent and -independent functions of TET1 to enable the activation of a critical set of germline reprogramming-responsive genes involved in gamete generation and meiosis. Our results also reveal an unexpected role for TET1 in maintaining but not driving DNA demethylation in gonadal PGCs. Collectively, our work uncovers a fundamental biological role for gonadal germline reprogramming and identifies the epigenetic principles of the PGC-to-gonocyte transition that will help to guide attempts to recapitulate complete gametogenesis *in vitro*.

In order to address the potential role and underlying molecular mechanisms of gonadal germline reprogramming (Fig. 1a), we first set out to investigate the dynamics of, and relationship between, 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC), as 5hmC has previously been implicated in DNA demethylation in PGCs^{3,6,9–11}. We did this quantitatively, and at single-base resolution, using liquid chromatography coupled with mass spectrometry (LC–MS/MS) alongside whole-genome bisulfite sequencing (WGBS) (Extended Data Fig. 1a) and AbaSI endonuclease digestion coupled with sequencing (Aba-seq)¹⁵ (Extended Data Fig. 1b–e). WGBS provides information regarding combined levels of 5mC and 5hmC¹⁶, whereas Aba-seq¹⁵ enables robust site-specific quantification and accurate genome-wide comparison of 5hmC levels within a given sample and between samples when combined with LC–MS/MS (see Methods, Extended Data Fig. 1b–e).

Using LC–MS/MS, we observed that global levels of genomic 5mC remain stable between migratory (E9.5) and early gonadal (E10.5) PGCs, whereas there was a significant reduction in the levels of 5mC between E10.5 and E11.5 and much more limited DNA demethylation between E11.5 and E13.5 (Fig. 1b). With respect to 5hmC, LC–MS/MS analysis revealed that global levels in PGCs are lower than those in mouse embryonic stem (ES) cells grown in serum-containing culture conditions (Fig. 1b). Furthermore, the global levels of 5hmC in PGCs are relatively constant between E9.5 and E13.5, with a slight decrease in females starting at E12.5 (Fig. 1b). Notably, levels of 5hmC

are consistently an order of magnitude lower than either total levels of 5mC at E10.5 or the amount of 5mC lost between E10.5 and E11.5 (Fig. 1b, c), demonstrating that global DNA demethylation is not accompanied by a reciprocal increase in levels of 5hmC, as has previously been suggested^{3,17}.

Consistent with our LC–MS/MS measurements, WGBS analysis revealed near-complete loss of combined 5mC and 5hmC between E10.5 and E11.5 at features within uniquely mapped regions of the genome, with limited amounts of further DNA demethylation observed between E11.5 and E12.5 (Extended Data Fig. 3a). Loss of DNA methylation was also observed at consensus repeat sequences, although some repetitive elements such as long interspersed nuclear element (LINE)-1A and endogenous retrovirus-intracisternal A particle (ERV-IAP) retrotransposons retained comparatively high combined levels of 5mC and 5hmC in E12.5 PGCs, as previously described⁸ (Extended Data Fig. 3b). Detailed analysis of 5hmC localization by Aba-seq in E10.5 PGCs revealed that, although global levels are lower (Fig. 1b), 5hmC localization in PGCs is remarkably similar to that of serum-grown mouse ES cells, even at imprint control regions (ICRs) (Extended Data Fig. 2a, b, f). Overall, 5hmC was enriched at putative active enhancers, present in intergenic regions and gene bodies, depleted at promoters, and absent on the vast majority of CpG islands (Extended Data Fig. 2b–f). With respect to transcription (Supplementary Table 7), both 5mC and 5hmC at promoter regions show an inverse relationship with gene expression levels (Extended Data Fig. 2c). Within gene bodies, 5mC and 5hmC are clearly enriched at expressed genes when compared to genes without detectable expression. A nonlinear relationship with gene expression is observed for 5hmC, whereas the combined levels of 5mC and 5hmC show a clear positive correlation with gene expression (Extended Data Fig. 2c).

Detailed analysis of 5hmC patterns across examined developmental stages showed that the majority of 5hmC is lost from uniquely mapped regions of the genome and re-localized to repetitive elements (Fig. 1d, Extended Data Fig. 3a, b). This relocalization was also clearly evident by immunofluorescence staining (Fig. 1e). Our data thus show that during reprogramming both 5mC and 5hmC are lost in PGCs throughout the uniquely mapped regions of the genome, although levels of 5hmC show a more gradual decrease (Extended Data Fig. 4b). However, this is not consistent with passive dilution of 5hmC through cell division³, as demonstrated by poor Pearson and Spearman's correlations between stages (Extended Data Fig. 4a, 5a). To the contrary, we conclude that 5hmC is a dynamic mark in PGCs.

We next explored the relationship between 5hmC deposition and DNA demethylation in gonadal PGCs between E10.5 and E12.5 for all initially methylated 2-kb windows (that is, windows with a minimum of 20% methylation at E10.5). DNA demethylation involving a 5hmC intermediate predicts a direct correlation between 5hmC appearance

¹MRC London Institute of Medical Sciences (LMS), Du Cane Road, London W12 0NN, UK. ²Institute of Clinical Sciences (ICS), Faculty of Medicine, Imperial College London, Du Cane Road, London W12 0NN, UK. ³New England Biolabs, Inc., 240 County Road, Ipswich, Massachusetts 01938, USA. [†]Present addresses: Research Institute of Molecular Pathology (IMP), Vienna Biocenter (VBC), Dr Bohr-Gasse 7, 1030 Vienna, Austria (V.H.); Medical Research Council Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK (H.B.).

*These authors contributed equally to this work.

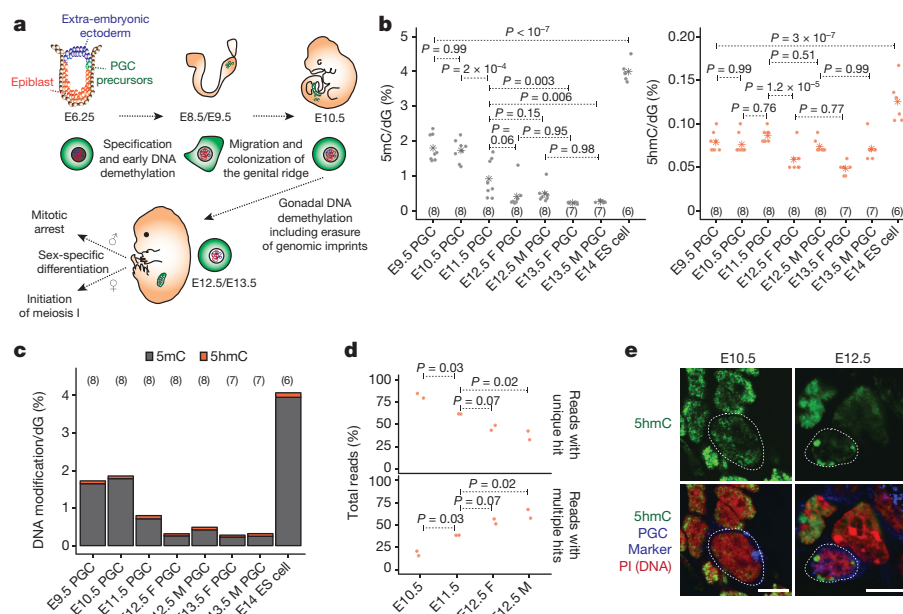


Figure 1 | 5mC and 5hmC dynamics during epigenetic reprogramming.

a, Key events during mouse PGC development. **b**, **c**, Individual 5mC (**b**, left), 5hmC (**b**, right) and combined 5mC and 5hmC (**c**) levels in mouse ES cells and E9.5 to E13.5 PGCs as determined by LC-MS/MS. Asterisks in **b** indicate mean values. Bars in **c** depict median values of biological replicates depicted in **b**. Sample numbers are indicated inside parentheses on each graph. Where relevant, 5mC and 5hmC level are quoted relative to total levels of deoxyguanine (dG). **d**, Redistribution of

5hmC from uniquely mapped parts of the genome to repetitive elements between E10.5 and E12.5. **b–d**, Data from males (M) and females (F) are shown separately for E12.5 and E13.5. **e**, Representative 5hmC immunostaining in E10.5 and E12.5 PGCs. Scale bars represent 10 μ m. PI, propidium iodide. **b**, **d**, *P* values are based on combined ANOVA and Tukey's post hoc test; *P* values in **b** are adjusted. Details for all figures regarding sample sizes and how samples were collected can be found in the 'Statistics and reproducibility' section of the Methods.

and 5mC loss (Extended Data Fig. 5a, b). However, we observed no correlation for E10.5 or E11.5 PGCs between either the total or relative levels of 5hmC and the extent to which the combined levels of 5mC and 5hmC decreased between these two stages (Extended Data Fig. 4c–f). Notably, for all initially methylated 2-kb windows, we did observe a negative correlation between the relative level of 5hmC and the combined levels of 5mC and 5hmC at E11.5 (Extended Data Fig. 4g). Thus, 5hmC represents a much higher proportion of the combined levels of 5mC and 5hmC at regions that are newly hypomethylated at E11.5, regardless of their original DNA methylation levels. Although 5hmC-depleted regions contain slightly more 5mC at E11.5 than regions enriched for 5hmC, sequences depleted of 5hmC in both E10.5 and E11.5 PGCs still undergo considerable DNA demethylation between these two stages (Extended Data Fig. 4h, i), indicating that the presence of detectable 5hmC is not a prerequisite for 5mC loss in gonadal PGCs. Our observations thus implicate involvement of 5hmC in the regulation of the locus-specific post-DNA-demethylation 5mC levels in germ cells rather than in the initial wave of global DNA demethylation (Extended Data Fig. 5c).

To expand on this observation, we used a previously published *Tet1*^{−/−} mouse model¹⁸ (Extended Data Fig. 6a–c). Initial LC-MS/MS analysis revealed that loss of TET1 leads to approximately 50% reduction in global levels of 5hmC in E10.5 *Tet1*^{−/−} germ cells (Fig. 2c). In agreement with the high level of TET1 expression at E12.5^{3,9,11} (Extended Data Fig. 6a–c), LC-MS/MS analysis also confirmed that TET1 represents the primary 5mC oxygenase in demethylated PGCs, with a decrease of approximately 85% in global levels of 5hmC observed in E14.5 *Tet1*^{−/−} germ cells (Fig. 2a, c). Importantly, the genomes of both *Tet1*^{−/−} and wild-type PGCs reached near-complete depletion of 5mC by E13.5 (Fig. 2b, d), highlighting that TET1-mediated 5mC oxidation is not directly responsible for the bulk of DNA demethylation in gonadal PGCs.

In support of our LC-MS/MS measurements, reduced representation bisulfite sequencing (RRBS) detected only a limited number of differentially methylated regions in E14.5 *Tet1*^{−/−} PGCs (Fig. 2e). Notably,

these regions initially undergo extensive DNA demethylation in both *Tet1*^{−/−} and wild-type PGCs, after which there is an increase in levels of 5mC between E12.5 and E14.5 specifically in *Tet1*^{−/−} PGCs (Fig. 2e). By contrast, levels of 5mC remain stable and/or undergo a slight further reduction between these stages in wild-type germ cells (Fig. 2e). The same DNA demethylation–remethylation kinetics were also observed at the few examples of previously reported^{9,10} germline gene promoters and ICRs that were found to be hypermethylated in E14.5 *Tet1*^{−/−} PGCs using RRBS (Extended Data Fig. 6d, e, Supplementary Table 8). Although considerable enrichment of 5mC is indeed observed at the *Dazl* promoter by targeted bisulfite sequencing in demethylated PGCs, the extent of hypermethylation observed at the *Peg3* and intergenic differentially methylated region (IG-DMR) ICRs is very limited (Extended Data Fig. 6f, g). Furthermore, for all three regions, very few clones showed full methylation and a number of clones had heterogeneous methylation patterns consistent with a stochastic failure to remove aberrant residual and/or *de novo* DNA methylation in *Tet1*^{−/−} PGCs (Extended Data Fig. 6f, g).

We next analysed the observed 5mC and 5hmC dynamics in combination with RNA sequencing (RNA-seq) datasets derived from E10.5–E14.5 PGCs (Extended Data Fig. 7). Initial clustering analysis of all genes on the basis of the dynamics of their promoter DNA methylation revealed that, although most promoters become completely demethylated, there is a small subset of transcriptionally silenced promoters that retain high levels of 5mC and 5hmC during global DNA demethylation (cluster 2; Extended Data Fig. 7a). These promoters overlap significantly with LINE1- and long terminal repeat (LTR)-containing endogenous retroviruses ($P = 9.5 \times 10^{-24}$ and $P = 7.2 \times 10^{-83}$, respectively; hypergeometric test) that are likely to determine this epigenetic status (Extended Data Fig. 3b). Overall, although high levels of 5mC and 5hmC at promoters are associated with transcriptional repression in E10.5 pre-reprogramming PGCs, loss of these marks does not generally result in transcriptional activation (Extended Data Fig. 7a).

As the influence of 5mC on the transcriptional activity of a gene has been shown in mammals to be highly dependent on promoter CpG

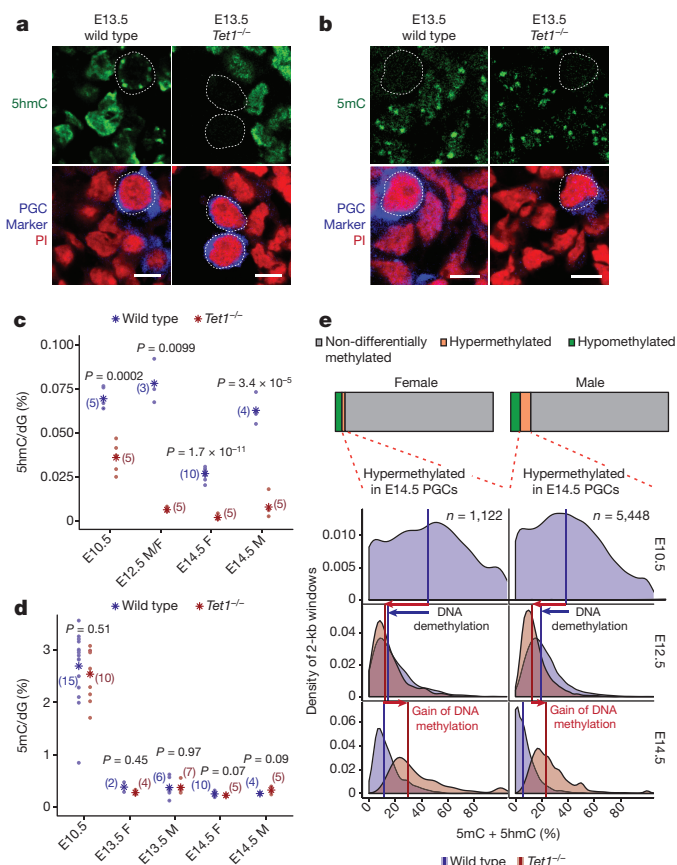


Figure 2 | TET1 safeguards but does not drive DNA demethylation.

a, b, Representative immunostaining of 5hmC (**a**) or 5mC (**b**) in E13.5 wild-type and *Tet1*^{-/-} PGCs. Scale bars represent 10 μ m. **c, d**, Global 5hmC (**c**) and 5mC (**d**) levels (assessed by LC-MS/MS) in wild-type and *Tet1*^{-/-} PGCs. Sample numbers are indicated in parentheses on each graph. Asterisks indicate mean values. *P* values are based on two-sided Student's *t*-tests. **e**, Top, proportion of differentially methylated regions in E14.5 *Tet1*^{-/-} PGCs (*P* < 0.05, greater than 10% methylation difference; *P* value derived from RnBeads software). Bottom, combined levels of 5mC and 5hmC (ascertained using RRBS) in E12.5 (middle plots) and E14.5 (bottom plots) *Tet1*^{-/-} (red) and wild-type (blue) PGCs for all E14.5 hypermethylated 2-kb windows. DNA modification levels from E10.5 wild-type PGCs are also shown (top plot). Median combined levels of 5mC and 5hmC are denoted by vertical lines.

content¹⁹, we performed clustering analysis specifically at genes with either high-CpG (HCPs), intermediate-CpG (ICPs) or low-CpG (LCPs) promoters¹⁹ (Fig. 3a and Extended Data Fig. 7b, c). Notably, this yielded a group of HCP genes that became demethylated during the course of germline epigenetic reprogramming, and showed progressive transcriptional activation (cluster 3; Fig. 3a). Differential expression analysis confirmed that these genes are significantly enriched among all genes that are upregulated concurrently with epigenetic reprogramming in PGCs (*P* < 0.001, hypergeometric test), with 45 genes commonly activated in both sexes (Fig. 3a–c). Considering their promoter methylation dynamics and the timing of their activation, we termed these 45 genes germline reprogramming-responsive (GRR) genes (Fig. 3c). Notably, GRR genes show significant enrichment for factors involved in gamete generation and meiosis, including *Dazl*, *Sycp1–Sycp3*, *Mael*, *Hormad1*, and *Rad51c* (Fig. 3c and Supplementary Tables 5 and 7).

Considering that GRR genes (*n* = 45) constituted less than 25% of the entire subset of HCP genes that undergo DNA demethylation (*n* = 226; Fig. 3a–c), DNA demethylation is probably an important factor for transcriptional activation of methylated HCPs, although other factors are also likely to be necessary. Indeed, GRR gene promoters showed both exceptionally high CpG density and levels of 5hmC compared to

other methylated and demethylating HCPs (Extended Data Fig. 9a, b). We also noted that, unusually for promoters, levels of 5hmC transiently increased at GRR gene promoters in PGCs immediately following the major wave of DNA demethylation (Extended Data Figs 3a, 9b). In addition, and in agreement with their high CpG density and levels of 5hmC^{20,21}, GRR gene promoters have been shown to be bound by TET1 in both mouse ES cells²¹ and PGCs⁹ (Fig. 3b).

The observed binding of TET1 is functionally relevant, as the extent of GRR gene upregulation is considerably lower in *Tet1*^{-/-} PGCs than in wild-type PGCs (Fig. 4a, Extended Data Fig. 9c, Supplementary Table 7). Although GRR gene promoters undergo normal DNA demethylation by E12.5 in the absence of TET1, they show slight hypermethylation later, in E14.5 *Tet1*^{-/-} PGCs (Fig. 4b). However, this limited DNA hypermethylation shows only weak correlation with the decreased expression (Extended Data Fig. 9e). Furthermore, lower expression of GRR genes in *Tet1*^{-/-} germ cells is already apparent at E12.5, in the absence of any methylation differences (Fig. 4a, b, Extended Data Fig. 9e), suggesting that TET1 is potentially acting as a transcriptional regulator outside its role in 5mC removal^{21,22}. In addition to GRR genes, transposable elements also become enriched in 5hmC during gonadal epigenetic reprogramming (Extended Data Figs 3b, 8). Alongside reduction in DNA methylation, some transposable elements show transcriptional activation concurrent with epigenetic reprogramming, especially from evolutionarily young retrotransposons (Extended Data Fig. 8, Supplementary Table 9). Notably, the lack of TET1 appears also to reduce the extent of transcriptional activation of transposable elements that are normally activated (Extended Data Fig. 8, Supplementary Table 9).

To further mechanistically probe the causal relationship between epigenetic reprogramming and GRR gene activation, we turned to an *in vitro* model. Serum-grown mouse ES cells represented an ideal system, as these cells are not germ line-restricted but have highly similar epigenetic modifications at GRR gene promoters to what is observed *in vivo* in pre-reprogramming gonadal PGCs (Extended Data Fig. 10a–d). Consistent with what we observed *in vivo*, promoter DNA demethylation represents a dominant epigenetic reprogramming event for GRR gene activation *in vitro*. *Dnmt1*^{-/-}*Dnmt3a*^{-/-}*Dnmt3b*^{-/-} triple knockout (DNMT TKO)²³ mouse ES cells display increased expression of GRR genes (Fig. 4c). However, even in the complete absence of DNA methylation, this is crucially dependent on the presence of TET1 as *Tet1*^{-/-} DNMT TKO mouse ES cells fail to activate GRR genes as a group (Fig. 4c, Extended Data Fig. 9f).

Although these *in vitro* observations clearly supported our *in vivo* data with respect to the roles of 5mC and TET1, the extent to which GRR genes were upregulated in DNMT TKO mouse ES cells (Fig. 4c) or in E10.5 PGCs that have undergone precocious DNA demethylation by conditional deletion of *Dnmt1* (*Dnmt1*^{CKO})²⁴ (Extended Data Fig. 9d) was relatively mild. We thus hypothesized that other factors, potentially including other epigenetic barriers, may regulate GRR gene expression. In this context, gonadal epigenetic reprogramming has previously been linked to the erasure of epigenetic information at various distinct levels^{4,25}, with removal of polycomb repressive complex 1 (PRC1) previously shown to coordinate the timing of meiosis initiation in DNA-demethylated E11.5/E12.5 PGCs²⁶. Remarkably, genes that are aberrantly upregulated following PRC1 deletion in PGCs show a significant enrichment for GRR genes (Extended Data Fig. 11a) and promoters of GRR genes in serum-grown mouse ES cells are enriched for RING1B binding and H2AK119ub (Extended Data Fig. 10a, e, f). In view of this, we simultaneously abolished both DNA methylation and PRC1 activity using highly specific chemical inhibition of PRC1 (using the inhibitor PRT4165²⁷) in DNMT TKO mouse ES cells to test the role of combined DNA methylation and PRC1 depletion on GRR gene regulation, thus mimicking gonadal epigenetic reprogramming. Culturing mouse ES cells with PRT4165 resulted in near complete inhibition of PRC1-mediated H2A ubiquitination after only 6 h of culture (Extended Data Fig. 11b). Notably,

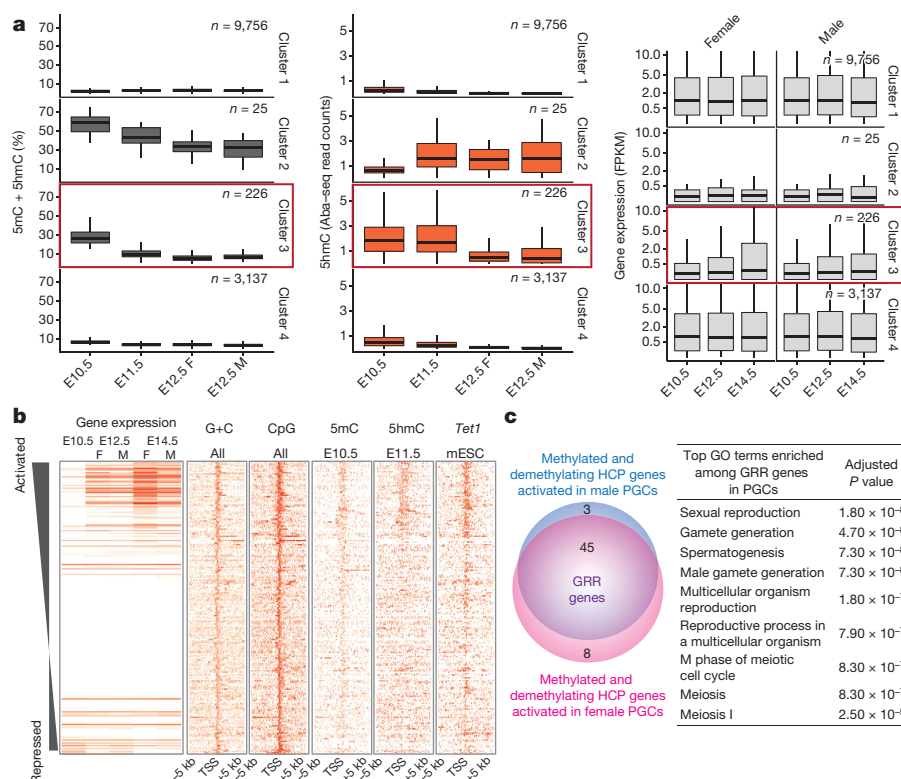


Figure 3 | GRR genes. **a**, Combined levels of 5mC and 5hmC at promoters determined by WGBS (left), levels of 5hmC at promoters ascertained by Aba-seq (centre), and gene expression levels (right) in consecutive stages of PGC development for HCP gene clusters (see Methods). For box plots, the upper and lower hinges correspond to the first and third quartiles, the centre line corresponds to the median, and the maxima and minima correspond to the highest and lowest value within $1.5 \times$ the inter-quartile range, respectively. FPKM, fragments per kilobase of transcript per million mapped reads. **b**, Genomic sequences centred on the transcription start site (TSS) of methylated and demethylating HCPs (cluster 3, Fig. 3a) ranked on the basis of the significance of upregulation between E10.5 and E14.5 in wild-type PGCs. Each horizontal line represents one gene; the intensity of red indicates the relative enrichment for the feature shown at the top of each column. The transcription start site ± 5 kb is shown in each column. mESC, mouse ES cell. **c**, Gene Ontology (GO) terms associated with GRR genes; adjusted *P* values are based on DAVID software.

combined inhibition of 5mC- and PRC1-mediated repression resulted in the activation of 33 out of 45 GRR genes whereas 25 and 10 genes were activated after the individual inhibition of either 5mC- or PRC1-mediated repression, respectively (Fig. 4d, Extended Data Fig. 11c). These observations demonstrate that gonadal epigenetic

reprogramming entails a composite erasure of epigenetic systems^{4,25} to potentiate the expression of GRR genes.

Our study has identified a set of GRR genes crucial for the correct progression of gametogenesis. These genes have unique promoter sequence characteristics, with high levels of both 5mC and 5hmC, and

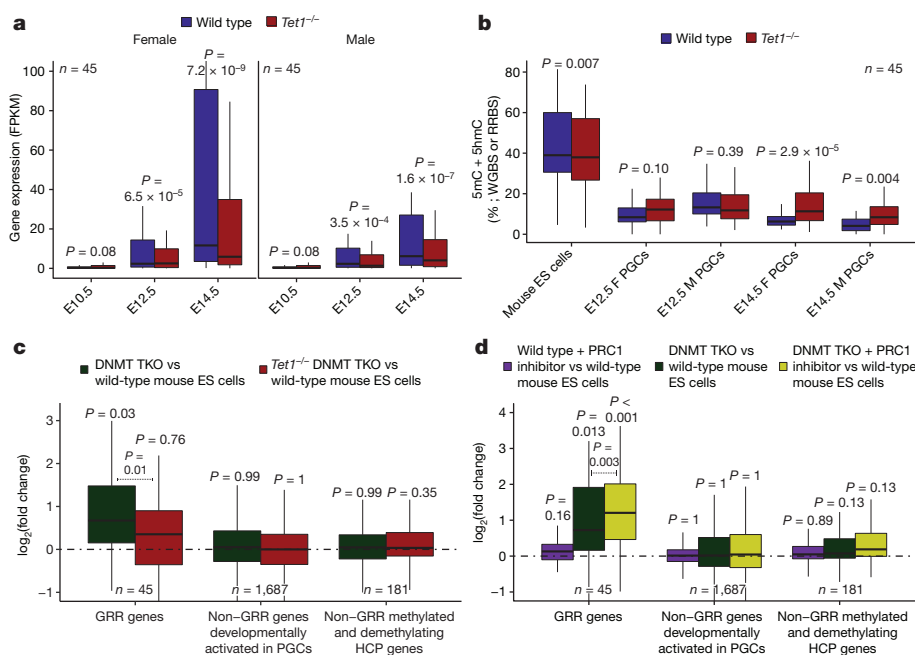


Figure 4 | Epigenetic principles of GRR gene activation. **a**, GRR gene expression dynamics in *Tet1*^{-/-} PGCs. *P* values are based on a two-sided paired Wilcoxon test. **b**, Combined levels of 5mC and 5hmC (ascertained using RRBS) at GRR genes in E12.5 or E14.5 *Tet1*^{-/-} (red) and wild-type (blue) PGCs. For comparison, combined levels of 5mC and 5hmC in mouse ES cells³⁰ (ascertained using WGBS) are shown. *P* values are based on paired two-sided Wilcoxon test. **c**, \log_2 -fold change between DNMT TKO (green) or *Tet1*^{-/-} DNMT TKO (red) and wild-type mouse ES cells (c) or between wild-type mouse ES cells after 6 h PRT4165 treatment

(purple), DNMT TKO after 6 h dimethylsulfoxide (DMSO) treatment (green) or DNMT TKO after 6 h PRT4165 treatment (yellow) and wild-type mouse ES cells after 6 h DMSO treatment (d) for GRR genes and other relevant genes sets. **c**, **d**, Family-wise error rate (FWER)-adjusted *P* values are based on gene set enrichment analysis (GSEA) software (see Methods for details). For all box plots, the upper and lower hinges correspond to the first and third quartiles, the centre line corresponds to the median, and the maxima and minima correspond to the highest and lowest value within $1.5 \times$ the inter-quartile range, respectively.

are targets of TET1 and PRC1. We show that the combined loss of DNA methylation and PRC1 repression is uniquely required for GRR gene activation, with this epigenetically poised state further requiring TET1 to potentiate both full and efficient activation. TET1 appears to be particularly important in female PGCs⁹, which initiate meiotic prophase soon after completion of epigenetic reprogramming, thus imposing a requirement for the timely expression of these genes. Although we observed slight hypermethylation at GRR gene promoters in E14.5 *Tet1*^{-/-} PGCs, our study also clearly shows that TET1 stimulates transcription of GRR genes via a DNA demethylation-independent mechanism^{21,22}. In this context, previous studies have shown that TET1 recruits OGT to gene promoters²², thus facilitating deposition of H3K4me3 via SET1–COMPASS²⁸, which leads to transcriptional activation. Furthermore, GRR gene promoters in mouse ES cells are marked by low but detectable H3K4me3, the levels of which are significantly decreased in the absence of TET1 without changes in DNA methylation (Fig. 4b, Extended Data Fig. 10g). Additionally, TET1 may also potentiate transcription through regulation of the levels of 5mC and 5hmC at non-promoter *cis*-active elements, such as enhancers. Last, but not least, our study shows that TET1 is not directly involved in initiation of global DNA demethylation during epigenetic reprogramming in gonadal PGCs; rather, we identify a critical role for TET1 in the subsequent removal of aberrant residual and/or *de novo* DNA methylation (Extended Data Fig. 12). This is reminiscent of the role of TET3-driven 5mC oxidation in protection against *de novo* DNA methylation during zygotic DNA demethylation²⁹, suggesting that global reprogramming events require efficient protection from *de novo* DNA methylation to stabilize the newly acquired epigenetic state after the removal of 5mC. Collectively, our results reinforce the idea that gonadal epigenetic reprogramming entails complex erasure of epigenetic information⁴ and suggest that a central function of this process is to ensure the timely and efficient activation of GRR genes, thus enabling progression towards gametogenesis (Extended Data Fig. 12).

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 21 July 2017; accepted 1 February 2018.

Published online 7 March 2018.

1. Lesch, B. J. & Page, D. C. Genetics of germ cell development. *Nat. Rev. Genet.* **13**, 781–794 (2012).
2. Guibert, S., Forné, T. & Weber, M. Global profiling of DNA methylation erasure in mouse primordial germ cells. *Genome Res.* **22**, 633–641 (2012).
3. Hackett, J. A. *et al.* Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science* **339**, 448–452 (2013).
4. Hajkova, P. *et al.* Chromatin dynamics during epigenetic reprogramming in the mouse germ line. *Nature* **452**, 877–881 (2008).
5. Hajkova, P. *et al.* Epigenetic reprogramming in mouse primordial germ cells. *Mech. Dev.* **117**, 15–23 (2002).
6. Hill, P. W., Amouroux, R. & Hajkova, P. DNA demethylation, Tet proteins and 5-hydroxymethylcytosine in epigenetic reprogramming: an emerging complex story. *Genomics* **104**, 324–333 (2014).
7. Lee, J. *et al.* Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells. *Development* **129**, 1807–1817 (2002).
8. Seisenberger, S. *et al.* The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol. Cell* **48**, 849–862 (2012).
9. Yamaguchi, S. *et al.* Tet1 controls meiosis by regulating meiotic gene expression. *Nature* **492**, 443–447 (2012).
10. Yamaguchi, S., Shen, L., Liu, Y., Sender, D. & Zhang, Y. Role of Tet1 in erasure of genomic imprinting. *Nature* **504**, 460–464 (2013).
11. Hajkova, P. *et al.* Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science* **329**, 78–82 (2010).
12. Hayashi, K. *et al.* Offspring from oocytes derived from *in vitro* primordial germ cell-like cells in mice. *Science* **338**, 971–975 (2012).

13. Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. & Saitou, M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* **146**, 519–532 (2011).
14. Hikabe, O. *et al.* Reconstitution *in vitro* of the entire cycle of the mouse female germ line. *Nature* **539**, 299–303 (2016).
15. Sun, Z. *et al.* High-resolution enzymatic mapping of genomic 5-hydroxymethylcytosine in mouse embryonic stem cells. *Cell Reports* **3**, 567–576 (2013).
16. Huang, Y. *et al.* The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS One* **5**, e8888 (2010).
17. Yamaguchi, S. *et al.* Dynamics of 5-methylcytosine and 5-hydroxymethylcytosine during germ cell reprogramming. *Cell Res.* **23**, 329–339 (2013).
18. Dawlaty, M. M. *et al.* Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell Stem Cell* **9**, 166–175 (2011).
19. Weber, M. *et al.* Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* **39**, 457–466 (2007).
20. Tahliliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
21. Williams, K. *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**, 343–348 (2011).
22. Vella, P. *et al.* Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. *Mol. Cell* **49**, 645–656 (2013).
23. Tsumura, A. *et al.* Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* **11**, 805–814 (2006).
24. Hargan-Calvopina, J. *et al.* Stage-specific demethylation in primordial germ cells safeguards against precocious differentiation. *Dev. Cell* **39**, 75–86 (2016).
25. Mansour, A. A. *et al.* The H3K27 demethylase Utx regulates somatic and germ cell epigenetic reprogramming. *Nature* **488**, 409–413 (2012).
26. Yokobayashi, S. *et al.* PRC1 coordinates timing of sexual differentiation of female primordial germ cells. *Nature* **495**, 236–240 (2013).
27. Ismail, I. H., McDonald, D., Strickfaden, H., Xu, Z. & Hendzel, M. J. A small molecule inhibitor of polycomb repressive complex 1 inhibits ubiquitin signaling at DNA double-strand breaks. *J. Biol. Chem.* **288**, 26944–26954 (2013).
28. Deplus, R. *et al.* TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *EMBO J.* **32**, 645–655 (2013).
29. Amouroux, R. *et al.* De novo DNA methylation drives 5hmC accumulation in mouse zygotes. *Nat. Cell Biol.* **18**, 225–233 (2016).
30. Hon, G. C. *et al.* 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. *Mol. Cell* **56**, 286–297 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. Elliot and T. Adejumo for help with fluorescence activated cell sorting, L. Game for help with next-generation sequencing, F. Krueger for providing consensus repetitive element sequences, M. Woodberry, A. Cameron, and J. Glegola for mouse husbandry, T. Carell for a gift of isotopically labelled deoxynucleoside standards and the members of the Hajkova laboratory for discussions and revisions of the manuscript. Work in the Hajkova laboratory is supported by MRC funding (MC_US_A652_5PY70), the FP7 EpiGeneSys network and an ERC grant (ERC-CoG-648879–dynamicmodifications) to P.H. The laboratory of Y.Z. and S.P. is supported by grant 1R44GM096723-01A1. P.H. is a member of the EMBO Young Investigator Programme. P.W.S.H. is a recipient of an MRC PhD Studentship and MRC-targeted Doctoral Prize Fellowship from Imperial College London.

Author Contributions P.H. and P.W.S.H. conceived the study; P.W.S.H. performed the experiments and analysed the data; H.G.L. carried out mouse ES cell experiments with the help of M.B. and M.R.-T.; C.E.R. generated *Tet1*^{-/-} DNMT TKO mouse ES cell line with the help of H.B.; R.A. carried out LC-MS/MS experiments and analysed the data with the help of S.L.; J.T. made Aba-seq libraries with support from Y.Z.; computational analysis was carried out by P.W.S.H. with the help of Z.S., G.D., V.H., and B.L.; R.V. performed experiments; P.W.S.H. and P.H. wrote the manuscript with assistance from S.P. and B.L.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to P.H. (petra.hajkova@lms.mrc.ac.uk).

Reviewer Information *Nature* thanks Y. Matsui and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Mice. All animal experiments were carried out under and in accordance with a UK Home Office Project Licence in a Home Office-designated facility. Except for direct comparisons with *Tet1*^{-/-} PGCs, wild-type PGCs were isolated from embryos produced by crossing outbred MF1 females with mixed background GOF18ΔPE-EGFP (ref. 5) transgenic males. The sex of embryos from E12.5 onwards was determined by visual inspection of the gonads. For the study of *Tet1*^{-/-} PGCs, the *Tet1* knockout mouse strain (B6;129S4-*Tet1*^{tm1.1jac/J})¹⁸ was purchased from the Jackson Laboratory and bred onto the GOF18ΔPE-EGFP transgenic mouse line. Wild-type and *Tet1*^{-/-} PGCs were isolated from embryos produced from crosses between *Tet1*-heterozygous GOF18ΔPE-EGFP-homozygous females and males. For genotyping of embryos produced by crossing *Tet1*-heterozygous GOF18ΔPE-EGFP-homozygous males and females, PCR was always carried out twice using two different sets of primers (see below) to confirm deletion of exon 4. The sex of the embryos from E12.5 onwards was determined by visual inspection of gonads and additionally confirmed by PCR for *Sry*. In all cases, matings were timed in such a way that appearance of a vaginal plug at noon was defined as E0.5.

The following genotyping primers were used in this study: TCAGGGAGCTCATG GAGACTA (*Tet1* forward primer 1); AACTGATTCCTTCGTGCAG (*Tet1* forward primer 2); TTAAAGCATGGGTGGGAGTC (*Tet1* reverse primer); TTGTCTAGAGAGCATGGAGGGCCATGTCAA (*Sry* forward primer); CCACTCTCTGTGACACTTTAGCCCTCCGA (*Sry* reverse primer).

PGC isolation by flow cytometry. PGC isolation was carried out as previously described⁴. In brief, the embryonic trunk (E10.5) or genital ridge (E11.5–E14.5) was digested at 37 °C for 3 min using 0.05% trypsin–EDTA (1 ×) (Gibco) or TrypLE Express (Thermo Fisher Scientific). Enzymatic digestion was followed by neutralization with DMEM/F-12 (Gibco) containing 15% fetal bovine serum (Gibco) and manual dissociation by pipetting. Following centrifugation, cells were resuspended in DMEM/F-12 supplemented with hyaluronidase (300 μg ml⁻¹; Sigma), and a single cell suspension was generated by manual pipetting. Following centrifugation, cells were resuspended in ice-cold PBS supplemented with poly-vinyl alcohol (10 μg ml⁻¹) and EGTA (0.4 mg ml⁻¹, Sigma). GFP-positive cells were isolated using an Aria III (BD Bioscience) or Aria III (BD Bioscience) flow cytometer and sorted into ice-cold PBS supplemented with poly-vinyl alcohol (10 μg ml⁻¹) and EGTA (0.4 mg ml⁻¹, Sigma).

Generation of *Tet1*^{-/-} DNMT TKO mouse ES cells. The *Tet1*^{-/-} DNMT TKO mouse ES cell line was generated by CRISPR–Cas9-mediated genome editing. pX330 (Addgene, 42230) with a short guide RNA targeting *Tet1*³¹ (GGTGCTGTGTCAGGGAGCTCA) was co-transfected with a reporter GFP plasmid into 5 × 10⁶ DNMT TKO mouse ES cells²³ using Lipofectamine 3000. On the next day, GFP positive cells were sorted using fluorescence activated cell sorting (FACS) (BD Bioscience FACS Aria III) to a 96-well plate. Cells were cultured for a week and then frozen down before extraction of genomic DNA. Colonies were screened for mutations using a surveyor assay (Surveyor Mutation Detection Kit from IDT, and Taq DNA polymerase from Qiagen). Selected clones of *Tet1*^{-/-} DNMT TKO mouse ES cells were further analysed by genotype sequencing, which confirmed the presence of a frameshift mutation. Loss of *Tet1* was verified by RNA-seq and western blot. The following primers were used for genotype sequencing and surveyor assay: 5' TTGTCTCTCTCTGACTGC 3' and 5' TGATTGATCAAAATAGGCCTGC 3'.

Mouse ES cell culture. J1 (wild type), DNMT TKO²³ and *Tet1*^{-/-} DNMT TKO mouse ES cells were cultured in FCS/leukaemia inhibitory factor (LIF) medium without feeders on 0.1% gelatin. FCS/LIF medium consists of GMEM (Gibco) supplemented with 10% FCS, 0.1 mM MEM nonessential amino acids, 2 mM L-glutamine, 1 mM sodium pyruvate, 0.1 mM 2-mercaptoethanol and mouse LIF (ESGRO, Millipore). For inhibitor experiments, mouse ES cells were plated at a density of 1.5 × 10⁴ cells per cm² and left overnight. On the next morning, medium was exchanged for FCS/LIF medium containing either 50 μM PRC1 inhibitor PRT4165²⁷ or DMSO control and cells were pelleted at the indicated time for analysis.

Aba-seq library preparation. Total DNA was isolated from 10,000 sorted PGCs using the QIAamp DNA Micro Kit (Qiagen). Aba-seq libraries for 5hmC profiling were constructed as previously described¹⁵. In brief, genomic DNA was glucosylated and then digested using AbaSI enzyme (NEB). Biotinylated P1 adapters were ligated onto the AbaSI-digested DNA, which was then fragmented using a Covaris S2 sonicator (Covaris), following the manufacturer's instructions. Sheared P1-ligated DNA was then captured by mixing with Dynabeads MyOne Streptavidin C1 beads (Life Technologies) according to the manufacturer's specifications. End repair and dA-tailing were carried out on the beads by using the NEBNext End Repair Module (NEB) and the NEBNext dA-tailing Module (NEB) at 20 °C and 37 °C, respectively, for 30 min. P2 adapters were ligated to the random sheared ends of the dA-tailed DNA. Finally, the entire DNA was amplified using Phusion DNA polymerase (NEB) with the addition of 300 nM forward primer (PCR_I) and

300 nM reverse primers (PCR_IIpe) for 16 cycles. The libraries were purified using AMPure XP beads (Beckman–Coulter) and sequenced on the Illumina HiSeq 2000 instrument.

WGBS library preparation. Total DNA was isolated from 10,000 sorted PGCs using the QIAamp DNA Micro Kit (Qiagen). In some cases, unmethylated λ-phage DNA (Promega) was spiked in following DNA isolation to assess bisulfite conversion rate. DNA was fragmented using a Covaris S2 sonicator (Covaris), following the manufacturer's instructions. Libraries were prepared following the NEBNext Library Prep protocol, with methylated adaptors and the following modifications. First, bisulfite conversion was carried out after adaptor ligation using the Imprint Modification Kit (Sigma). Second, PCR enrichment was carried out for 16 cycles using the NEXTFlex Bisulfite-Seq Kit for Illumina Sequencing (Bioo Scientific) master mix and the NEBNext Library Prep universal and index primers (NEB). The libraries were purified using AMPure XP beads (Beckman–Coulter). Libraries were sequenced on the Illumina HiSeq 2000 or 2500 instrument.

RRBS library preparation. Total DNA from FACS-sorted PGCs isolated from individual *Tet1*^{-/-} or wild-type embryos was extracted using the ZR-Duet DNA-RNA MiniPrep kit (Zymo), and DNA from between two and five embryos (equivalent to 1,000 to 8,000 cells) of the same genotype, stage and sex was pooled and concentrated to 26 μl final volume using the Savant SpeedVac Concentrator (Thermo Fisher Scientific) according to the manufacturer's instructions. Genomic DNA was digested using 20 U of MspI enzyme (NEB) in NEB buffer 2 at 37 °C for 3 h, and digested DNA was purified using AMPure XP beads (Beckman–Coulter). Libraries were made following the NEBNext Ultra DNA Library Prep protocol with methylated adaptors and the following modifications. First, bisulfite conversion was carried out after adaptor ligation using the Imprint Modification Kit (Sigma). Second, PCR enrichment was carried out for 18 cycles using the KAPA Uracil⁺ DNA polymerase master mix (KAPA Biosystems) and the NEBNext Library Prep universal and index primers (NEB). The libraries were purified using AMPure XP beads (Beckman–Coulter). Pooled libraries were sequenced on the Illumina HiSeq 2500 instrument, using the 'dark sequencing' protocol, as previously described³².

RNA-seq library preparation. For the study of *Tet1*^{-/-} PGCs, total RNA from sorted PGCs isolated from individual *Tet1*^{-/-} or wild-type embryos was extracted using ZR-Duet DNA-RNA MiniPrep kit (Zymo), and RNA from between two and six embryos (equivalent to 1,000 to 8,000 cells) of the same genotype, stage and sex was pooled and concentrated to 6 μl final volume using the RNA Clean and Concentrator 5 kit (Zymo). For the study of wild-type PGCs isolated from embryos produced by crossing MF1 females with GOF18ΔPE-EGFP males, total RNA from 600–1,000 sorted E10.5 PGCs was isolated using the Nucleospin RNA XS kit (Macherey–Nagel). cDNA synthesis and amplification (15 cycles) was performed with the SMARTer Ultra Low Input RNA kit (Clontech) using between 100 pg and 3 ng total RNA, according to the manufacturer's instructions. The amplified cDNA was fragmented using a Covaris S2 sonicator (Covaris), following the manufacturer's instructions. Sheared cDNA was converted to sequencing libraries using the NEBNext DNA Library Prep kit (NEB), following the manufacturer's instructions and using 15 cycles of amplification. For the study of mouse ES cells, total RNA was isolated using ZR-Duet DNA-RNA MiniPrep kit (Zymo). cDNA synthesis and library preparation were performed starting with 500 ng total RNA using the NEBNext Ultra Library Prep Kit (NEB) and the NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB) according to the manufacturer's instructions. All libraries were purified by AMPure XP beads (Beckman–Coulter) and sequenced on the Illumina HiSeq 2500 instrument.

WGBS and Tet-assisted bisulfite sequencing (TAB-seq) alignment and downstream analysis. Raw reads were first trimmed using Trim Galore (v.0.3.1) with the '--paired' and '--trim1' options. Alignments were carried out to the mouse genome (mm9, NCBI build 37) using Bismark (v.0.13.0) with the '-n 1' parameter; where appropriate, the λ phage genome was added as an extra chromosome. Aligned reads were deduplicated with deduplicate_bismark. Where appropriate, the bisulfite conversion rate was computed using reads aligned to the λ phage genome and using the to-mr script ('-m bismark' parameters) and bsrscript ('-N' parameter) of Methpipe (v.3.3.1). CpG-methylation calls were extracted from the deduplicated mapping output using the Bismark methylation extractor. The number of methylated and unmethylated cytosines in a CpG context was extracted using bismark2bedGraph and coverage2cytosine. Symmetric CpGs were merged using a custom R script. For all downstream analysis, only symmetric CpGs with a minimum of 8 × coverage were used. All WGBS analysis was carried out on data from merged biological replicates. For assessing DNA modification levels at specific repetitive elements, Bismark (v.0.14.4) was used to map all reads from each dataset against consensus sequences constructed from Repbase with the '-n 1' parameter set. CpG-methylation calls were extracted from the mapping output using the Bismark methylation extractor (v.0.14.4).

The mapBed function of BEDtools (v.2.24.0) was used to compute the combined level of 5mC and 5hmC for the following genomic features: 1) all 2-kb windows

(containing a minimum of four symmetric CpGs); 2) gene promoters (defined as Ensembl 67 gene start sites $-1\text{ kb}/+500\text{ bp}$); 3) gene bodies (defined as the region contained within Ensembl 67 gene start and gene end sites); 4) putative active enhancers in day 6 PGC-like cells³³; 5) ICRs; 6) CpG islands (UCSC); 7) intergenic regions. For metagene plots, a genomic feature was divided into equally sized bins using BEDtools (v.2.24.0), including: 1) gene bodies (defined as the region contained within Ensembl 67 gene start and gene end sites) $\pm 0.5 \times$ gene body length (100 bins); 2) putative active enhancers in day 6 PGC-like cells³³ $\pm 1 \times$ putative active enhancer length (90 bins); and 3) CpG islands (UCSC) $\pm 1 \times$ CpG island length (90 bins). In all cases, the combined level of 5mC and 5hmC was expressed as the mean of individual CpG sites.

For *k*-means clustering of the combined mean levels of 5mC and 5hmC, HCPs, ICPs and LCPs were defined using previously published parameters^{19,34}. In brief, LCPs contain no 500-bp window with a CpG ratio > 0.45 ; HCPs contain at least one 500-bp window with a CpG ratio > 0.65 and GC content $> 55\%$; ICPs do not meet the previous criteria.

To determine locus-specific methylation levels in wild-type mouse ES cells grown in serum-containing medium, raw WGBS reads were downloaded from GSE48519³⁰ and processed as above. TAB-seq reads for E14 mouse ES cells were downloaded from GSE36173³⁵ and processed as above, with the exception that only symmetric CpGs with a minimum of $12\times$ coverage were used.

Aba-seq alignment and downstream analysis. For the uniquely mapped part of the genome, Aba-seq reads were processed as previously described¹⁵. In brief, raw sequencing reads were trimmed to remove adaptor sequences and low quality bases using Trim Galore. The trimmed reads were mapped to the mouse genome (mm9, NCBI build 37) using Bowtie (v.0.12.8) with parameters ‘-n 1 -l 25 --best --strata -m 1’. Detection of 5hmC was based on the recognition sequence and cleavage pattern of the AbaSI enzyme (5' CN₁₁₋₁₃↓N₉₋₁₀G 3'/3' GN₉₋₁₀↓N₁₁₋₁₃C 5') using custom Perl scripts. For assessing the relative enrichment of 5hmC at repetitive elements and non-repetitive elements, Aba-seq alignments were divided into two groups: unique (single best alignment) and ambiguous (map to multiple locations with equal alignment score). Both groups were then mapped separately to the repetitive elements defined by the RepeatMasker track of mm9 (UCSC Genome Browser). For comparison with the levels of 5hmC in mouse ES cells, Aba-seq reads were downloaded from GSE42898¹⁵ and aligned in the same way.

To quantify the relative levels of 5hmC at symmetric CpGs in the uniquely mapped part of the genome, the number of counts per symmetric CpG for a given sample were normalized to the combined number of uniquely mapped and ambiguously mapped reads for a given library, and then further multiplied by a stage-specific normalization factor that was based on the mean level of 5hmC for each stage as computed by LC-MS/MS (E14 ESC = 1.64; E10.5 = 1.0; E11.5 = 1.13; E12.5F = 0.76; E12.5M = 1.0). All symmetric CpGs falling within genomic intervals blacklisted by the mouse (mm9) ENCODE project were excluded from all further downstream analysis. Unless stated otherwise, all Aba-seq analysis was carried out on data from merged biological replicates.

The mapBed function of BEDtools (v.2.24.0) was used to compute the level of 5hmC for the same genomic features as carried out with WGBS datasets (see above). In all cases, the level of 5hmC was expressed as the mean of individual CpG sites.

To identify 5hmC-enriched or -depleted regions in E10.5 and E11.5 PGCs, the mm9 genome was first divided into 2-kb windows (minimum four symmetric CpGs) and the mean level of 5hmC for each window was computed using BEDtools (v.2.24.0). To determine the significance of 5hmC enrichment in each 2-kb window, upper-tail (to determine 5hmC enriched regions) or lower-tail (to determine 5hmC depleted regions) Poisson probability *P* values were computed using $\text{ppois}(x, \lambda)$, where *x* is the observed 5hmC mean value for each 2-kb window and λ is the mean of 5hmC mean values for all 2-kb windows at E10.5. Benjamini-Hochberg correction was then applied to correct for multiple testing, giving a final adjusted upper-tail and lower-tail *P* value for each 2-kb window. Windows with an adjusted upper-tail *P* < 0.05 were considered relatively enriched for 5hmC and windows with adjusted lower-tail *P* < 0.05 were considered relatively depleted for 5hmC.

For assessing relative enrichment of 5hmC at specific repetitive elements, Bowtie was used to map all reads from each dataset against consensus sequences constructed from Repbase with parameters ‘-n 1 -M 1 --strata --best’. The number of reads mapped to each sequence within a given sample was first normalized to the library size of that particular sample, and then normalized to both a stage-specific normalization factor that was based on the mean level of 5hmC for each stage as computed by LC-MS/MS (E10.5 = 1.0; E11.5 = 1.13; E12.5F = 0.76; E12.5M = 1.0) and the mean proportion of reads mapped to a given sequence in E10.5 PGCs.

RRBS alignment and downstream analysis. Raw RRBS reads were first trimmed using Trim Galore (v.0.3.1) with the ‘-rrbs’ parameter. Alignments were carried out to the mouse genome (mm9, NCBI build 37) using Bismark (v.0.13.0) with the

‘-n 1’ parameter. CpG-methylation calls were extracted from the mapping output using the Bismark methylation extractor (v.0.13.0). The number of methylated and unmethylated cytosines in a CpG context was extracted using bismark2bedGraph.

RnBeads (v.1.0.0) and RnBeads.mm9 (v.0.99.0) (‘filtering.missing.value.quantile’ set to 0.95 and ‘filtering.missing.coverage.threshold’ set to 8) were used to identify differentially methylated regions between the two test groups for the following genomic features: 1) all 2-kb windows (containing a minimum of four symmetric CpGs); 2) gene promoters (defined as Ensembl 67 gene start sites $-1\text{ kb}/+500\text{ bp}$); and 3) ICRs (mm9 genome). The following data were extracted from the output of RnBeads: 1) the mean methylation level for each group (that is, stage, sex and/or genotype) for each commonly covered test region; 2) the difference in methylation means between two groups for each commonly covered test region; and 3) the *P* value representing the significance of the difference in methylation means between two groups for each commonly covered test region. Differentially methylated regions were identified as regions with *P* < 0.05 and a difference in methylation means between two groups greater than 10%.

For assessing DNA modification levels at specific repetitive elements, Bismark (v.0.14.4) was used to map all reads from each dataset against consensus sequences constructed from Repbase with the ‘-n 1’ parameter set. CpG-methylation calls were extracted from the mapping output using the Bismark methylation extractor (v.0.14.4). The number of methylated and unmethylated cytosines in a CpG context were extracted using bismark2bedGraph and coverage2cytosine. Differentially methylated consensus repeats were identified as regions with a *P* < 0.05 (as computed by two-sided Student's *t*-test) and a difference in methylation means between two groups greater than 10%.

Hydroxymethylated-DNA immunoprecipitation alignment and downstream analysis. Raw hydroxymethylated-DNA immunoprecipitation (hMeDIP) sequencing and input reads for E14 mouse ES cells were downloaded from GSE28500³⁶ and aligned to the mouse genome (mm9, NCBI build 37) using Bowtie (v.0.12.8) with parameters ‘-n 2 -l 25 -m 1’. BEDtools multicov was used to identify the number of hMeDIP and input reads overlapping each 2-kb window (containing a minimum of four symmetric CpGs). Final levels of 5hmC for each 2-kb window were determined by first normalizing the number of overlapping hMeDIP reads (normalized to library size) by the number of overlapping input reads (normalized to library size) and then dividing this value by the number of symmetric CpGs contained within the 2-kb window.

Chromatin immunoprecipitation sequencing alignment and downstream analysis. For putative active enhancer calling, raw chromatin immunoprecipitation sequencing (ChIP-seq) reads for H3K4me3, H3K27me3 and H3K27Ac in day 6 PGC-like cells were downloaded from GSE60204³³ and raw ChIP-seq reads for H3K4me3, H3K27me3, H3K4me1 and H3K27Ac in wild-type mouse ES cells were downloaded from GSE48519³⁰. Reads were aligned to the mouse genome (mm9, NCBI build 37) with Bowtie (v.0.12.8 or v.1.0.0) with parameters ‘-n 2 -l 25 -m 1’ and ‘-C’ where appropriate. Subsequent ChIP-seq analysis was carried out on data from merged biological replicates. To identify putative active enhancers, we first generated an eight-state chromatin model using ChromHMM. Putative active enhancers were defined as all regions not overlapping any potential promoter regions (Ensembl 67 gene start sites $-1\text{ kb}/+500\text{ bp}$) and contained within the (H3K27Ac⁺/H3K4me3⁻/H3K27me3⁻) chromatin state in day 6 PGC-like cells or (H3K4me1⁺/H3K27Ac⁺/H3K4me3⁻/H3K27me3⁻) in wild-type mouse ES cells.

For analysis of epigenetic modifications and modifiers around transcription start sites (Ensembl 67): raw ChIP-seq reads for: TET1 binding in wild-type serum-grown mouse ES cells were downloaded from GSE24843²¹; H2AK119ub1 levels in wild-type serum-grown mouse ES cells were downloaded from GSE34520³⁷; RING1B binding in wild-type serum-grown mouse ES cells were downloaded from ERP005575³⁸; and for H3K4me3 in wild-type and *Tet1*^{-/-} serum-grown mouse ES cells were downloaded from GSE48519³⁰. Reads were aligned to the mouse genome (mm9, NCBI build 37) with Bowtie (v.0.12.8 or v.1.0.0) with parameters ‘-n 2 -l 25 -m 1’. Subsequent ChIP-seq analysis was carried out on data from merged biological replicates. For computing the ChIP-seq signal around transcription start sites, the genomic interval around the Ensembl 67 gene transcription start sites ($\pm 5\text{ kb}$ or 2 kb) was divided into 100 (or 40) equally sized bins using BEDtools makewindows. BEDtools multicov was then used to compute the number of test and control reads overlapping each bin. The total numbers of test and control reads per bin for each sample were normalized to the appropriate library size, and fold enrichment for each bin was determined by dividing the number of normalized ChIP-seq test sample reads by the number of normalized ChIP-seq control sample reads. For computing ChIP-seq signal at gene promoters, the genomic interval around the Ensembl 67 gene start sites $+500\text{ bp}/-1\text{ kb}$ was used to compute the number of test and control reads overlapping each region.

RNA-seq alignment and downstream analysis. For the study of *Tet1*^{-/-} and wild-type *Tet1* PGCs, Illumina and SMART-seq adapters from the sequencing reads were first trimmed using Trimmomatic. For other RNA-seq libraries, fastq

files generated from output of next generation sequencing were used directly for alignment. RNA-seq reads were aligned to the mouse genome (mm9, NCBI build 37) with Bowtie (v.0.12.8) and Tophat (v.2.0.2) with options '-N 2 --b2-very-sensitive --b2-L 25'. Annotations from Ensembl Gene version 67 were used as the gene model with Tophat. Read counts per annotated gene were computed using HTSeq (v.0.5.3p9) and the expression of each gene was quantified by computing the FPKM using a custom R script. Genes were assigned to an expression-level bin based on the mean FPKM values of the two biological replicates. Differential expression analysis was performed using DESeq2 (v.1.6.3), and genes with an adjusted $P < 0.05$ were considered to be differentially expressed. For determining gene expression levels in wild-type and *Dnmt1*^{CKO} and matched wild-type E10.5 PGCs, raw RNA-seq reads were downloaded from GSE74938²⁴ and processed as described above.

HCPs that were methylated and demethylated in PGCs during epigenetic reprogramming (cluster 3, Fig. 3a) were ranked on the basis of the significance of activation (α) between gene expression in E10.5 and E14.5 PGCs (Fig. 3b). In the case where β represents the directionality of fold change (that is, if $\log_2(\text{fold change}) < 0$, $\beta = -1$; else $\beta = 1$) and γ represents the adjusted P value as computed by DESeq2, $\alpha = \beta \times (1 - \gamma)$. For comparing expression levels of the GRR gene set in 1) wild-type, DNMT TKO, and *Tet1*^{-/-} DNMT TKO mouse ES cells (Fig. 4c); 2) in wild-type + 6 h DMSO treatment, DNMT TKO + 6 h DMSO treatment, wild-type + 6 h PRT4165 treatment, DNMT TKO + 6 h PRT4165 treatment (Fig. 4d); 3) *Tet1*^{-/-} E14.5 PGCs against wild-type E14.5 PGCs (Extended Data Fig. 9c); or 3) *Dnmt1*^{CKO} E10.5 PGCs against wild-type E10.5 PGCs (Extended Data Fig. 9d) pairwise differential expression analysis was initially carried out by DESeq2 for each condition against each other condition. For each pairwise differential expression test, each gene was assigned a statistic α , where if β represents the $\log_2(\text{fold change})$ and γ represents the adjusted P value as computed by DESeq2, $\alpha = \beta \times (1 - \gamma)$. The gene list ranked on the basis of α was subsequently used for GSEA for testing general up- or downregulation of the combined GRR gene sets and GSEA hallmark gene sets. GSEA FWER-adjusted P values were subsequently used. For overlap between GRR genes and genes repressed by PRC1 in PGCs (Extended Data Fig. 11a), the list of genes labelled as upregulated in E11.5 and/or E12.5 PRC1 conditional knockout (PRC1 CKO) PGCs was downloaded from ref. 26.

For classification of GRR genes (Extended Data Fig. 10, Supplementary Table 5), pairwise differential expression analysis was first carried out. 5mC-reprogramming-dependent GRR genes were defined as genes that were: 1) upregulated in DNMT TKO versus wild type, DNMT TKO + PRC1 inhibitor versus wild type, and DNMT TKO + PRC1 inhibitor versus wild type + PRC1 inhibitor; and 2) not upregulated in wild type + PRC1 inhibitor versus wild type. PRC1-reprogramming-dependent GRR genes were defined as genes that were: 1) upregulated in wild type + PRC1 inhibitor versus wild type, DNMT TKO + PRC1 inhibitor versus wild type, and DNMT TKO + PRC1 inhibitor versus DNMT TKO; and 2) not upregulated in DNMT TKO versus wild type. 5mC/PRC1-reprogramming-dependent GRR genes were defined as genes that were either: 1) upregulated in wild type + PRC1 inhibitor versus wild type, DNMT TKO versus wild type, DNMT TKO + PRC1 inhibitor versus wild type, DNMT TKO + PRC1 inhibitor versus DNMT TKO, and DNMT TKO + PRC1 inhibitor versus wild type + PRC1 inhibitor; or 2) upregulated in DNMT TKO + PRC1 inhibitor versus wild type, DNMT TKO + PRC1 inhibitor versus DNMT TKO, and DNMT TKO + PRC1 inhibitor versus wild type + PRC1 inhibitor, and not upregulated in wild type + PRC1 inhibitor versus wild type and DNMT TKO versus wild type. 5mC/PRC1-reprogramming-independent or -insufficient GRR genes were defined as genes that were not upregulated in DNMT TKO versus wild type, DNMT TKO + PRC1 inhibitor versus wild type, DNMT TKO + PRC1 inhibitor versus wild type + PRC1 inhibitor, and wild type + PRC1 inhibitor versus wild type. Genes that did not fall into one of these four classes were described as low confidence classification genes.

Detection of TET1, 5mC and 5hmC by immunofluorescence. The embryonic trunk (E10.5) or genital ridge (E12.5/E13.5) was first fixed in 2% PFA (in PBS) for 30 min at 4°C. Following fixation, tissue was washed in PBS three times for 10 min and then incubated in 15% sucrose in PBS overnight. After rinsing with 1% BSA in PBS the following day, the tissue was embedded in OCT Embedding Matrix (Thermo Fisher Scientific Raymond Lamb) and frozen using liquid nitrogen. Samples were subsequently stored at -80°C. A Leica CM 1950 cryostat was used to cut 10 µm sections from the frozen embedded tissue. Sections were settled on poly-lysine slides (Thermo Fisher Scientific) and post-fixed with 2% PFA in PBS for 3 min.

For detection of TET1, sections were washed three times for 5 min with PBS. After incubating for 30 min at room temperature in 1% BSA in PBS containing 0.1% Triton X-100, the sections were incubated with the listed primary antibodies at 4°C overnight in the same buffer. Sections were subsequently washed three times in 1% BSA in PBS containing 0.1% Triton X-100 for 5 min and incubated with

secondary antibodies in the same buffer for 1 h in the dark at room temperature. After secondary antibody incubation samples were washed three times in PBS for 5 min. DNA was then stained with DAPI (100 ng ml⁻¹). After a final wash in PBS for 10 min, the sections were mounted with Vectashield (Vector Laboratories).

For detection of 5hmC and 5mC, sections were washed three times for 5 min with PBS. Post-fixed sections were first permeabilized for 30 min with 0.5% Triton X-100 (in 1% BSA in PBS), and subsequently treated with RNase A (10 mg ml⁻¹; Roche) in 1% BSA in PBS for 1 h at 37°C. After three 5 min washes with PBS, sections were incubated with 4N HCl for 10–20 min at 37°C to denature genomic DNA and then washed three times for 10 min with PBS. After incubating for 30 min at room temperature in 1% BSA in PBS containing 0.1% Triton X-100, the sections were incubated with the listed primary antibodies at 4°C overnight in the same buffer. Sections were subsequently washed three times in 1% BSA in PBS containing 0.1% Triton X-100 for 5 min and incubated with secondary antibodies in the same buffer for 1 h in the dark at room temperature. After secondary antibody incubation samples were washed three times in PBS for 5 min. DNA was then stained with propidium iodide (0.25 mg ml⁻¹). After a final wash in PBS for 10 min, the sections were mounted with Vectashield (Vector Laboratories).

The following primary antibodies were used in this study: anti-SSEA1 (gift from P. Beverly via G. Durcova Hills); anti-MVH (Abcam 27591 or Abcam 13840); anti-5hmC (Active motif 39791), anti-5mC (Diagenode C15200081-100); anti-TET1 (GeneTex GTX125888); anti-GFP (Abcam 5450). The following secondary antibodies were used in this study: Alexa Fluor 647 goat anti-mouse IgM (Invitrogen A21238); Alexa Fluor 488 goat anti-rabbit IgG (Invitrogen A11008); Alexa Fluor 405 goat anti-mouse IgG 1:300 (Invitrogen A31553); Alexa Fluor 488 goat anti-mouse IgG 1:300 (Invitrogen A11001); Alexa Fluor 405 goat anti-rabbit IgG 1:300 (Invitrogen A31556); Alexa Fluor 568 donkey anti-rabbit IgG (Invitrogen A10042); Alexa Fluor 488 donkey anti-goat IgG (Invitrogen A11055).

Locus-specific bisulfite sequencing. Bisulfite treatment of genomic DNA was carried out using the Imprint DNA modification kit (Sigma). The following primers were used for the semi-nested amplification of the *Dazl* promoter: forward 1: GATTT TTGTTATTTTTAGTTTCTTTTAGGAT; forward 2: TTTATTTAAGTTATTAT TTTAAAAATGGTATT; reverse: AGAAACAAGCTAGGCCAGCTGAGAG AATTCT. The following primers were used for the semi-nested amplification of the IG-DMR ICR: forward 1: GTGTTAAGGTATATTATGTAGTGTAGG; forward 2: ATATTATGTAGTGTAGGTAAGGATTGTG; reverse: TACAACC CTTCCCTCACTCCAAAAATT. The following primers were used for the nested amplification of the *Peg3* ICR: forward 1: TTTTATAGATTTTGTGTTGGGGG TTTTAAATA; forward 2: TTGATAATAGTAGTTTGATTGGTAGGGTGT; reverse 1: AATCCCTATCACCTAAATAACATCCCTACA; reverse 2: ATCTACAACCTTATCAATTACCCTTAAAAA. Methylation levels were assessed using QUMA, applying default settings with duplicate bisulfite sequences excluded.

Mass spectrometry. Genomic DNA from between 100 and 2,000 FACS-sorted PGCs was extracted using ZR-Duet DNA/RNA Miniprep kit (Zymo Research) following the manufacturer's instructions and eluted in LC-MS grade water. DNA was digested to nucleosides using a digestion enzyme mix provided by NEB. A dilution series made with known amounts of synthetic nucleosides and the digested DNA was spiked with a similar amount of isotope-labelled nucleosides (provided by T. Carell) and separated on an Agilent RRHD Eclipse Plus C18 2.1 × 100 mm 1.8 µm column using the UHPLC 1290 system (Agilent) and an Agilent 6490 triple quadrupole mass spectrometer as previously described²⁹. To calculate the quantity of individual nucleosides, standard curves representing the ratio of unlabelled over isotope-labelled nucleosides were generated and used to convert the peak-area values to corresponding quantities. The threshold for quantification is a signal-to-noise above ten (calculated with a peak-to-peak method).

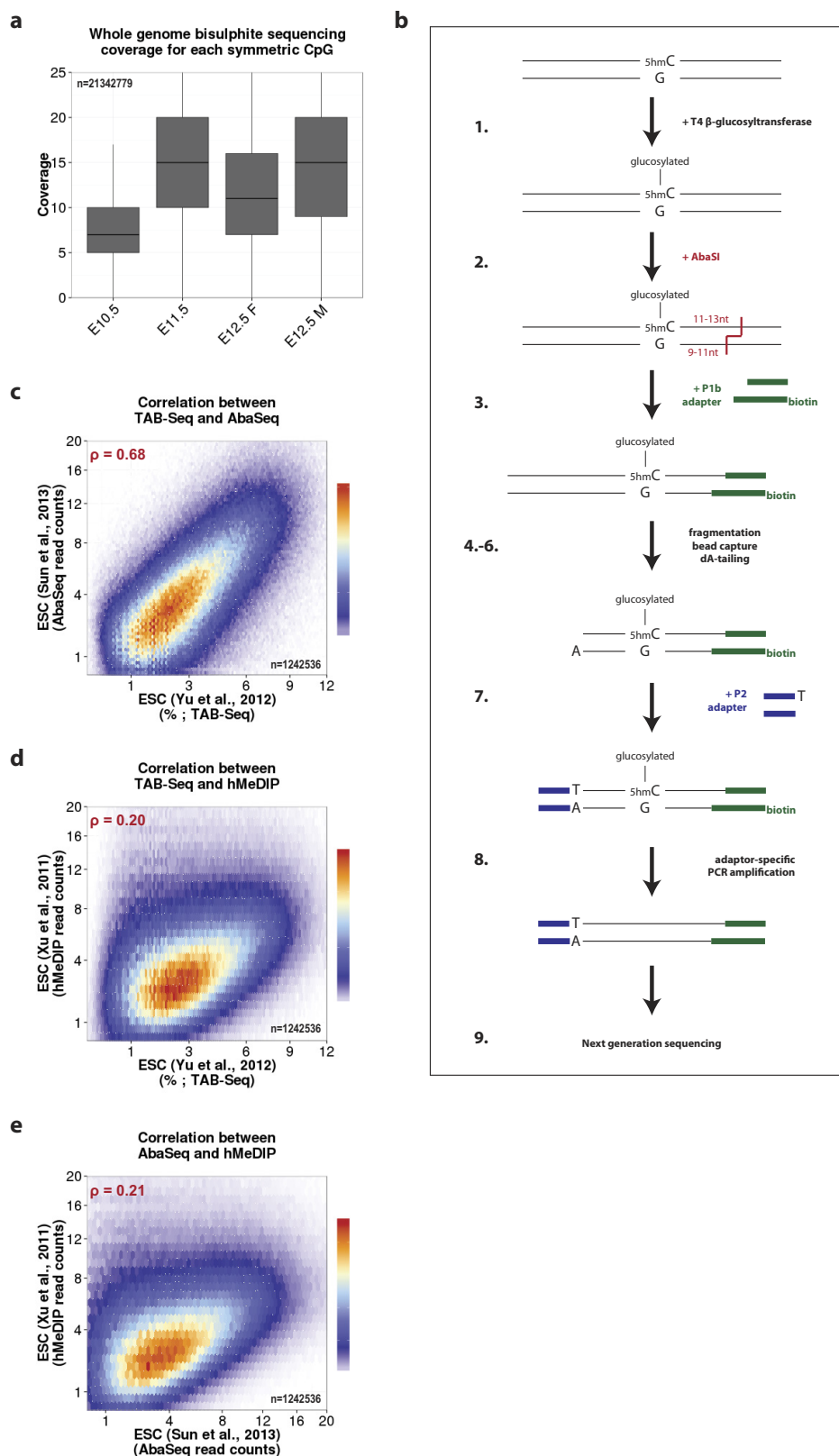
Western blotting. Mouse ES cells were lysed by sonication in RIPA buffer (150 mM sodium chloride, 1.0% Triton X-100, 0.5% sodium deoxychlorate, 0.1% sodium dodecylsulfate, 50 mM Tris pH 8.0) and protease-inhibitor cocktail (Roche, 11 697 498 001). Cell debris was removed by centrifugation at 14,000g for 5 min at 4°C. Protein levels were quantified using the BCA protein assay (Thermo Fisher Scientific, 23227). Each protein extract (2 µg for H2A and H2Aub or 20 µg for TET1) was loaded onto a 15% or 8% SDS polyacrylamide gel and transferred to a PVDF membrane after electrophoresis. Membranes were blocked with 5% BSA for 1 h and then incubated overnight at 4°C with primary antibodies at the following dilutions: anti-H2A antibody (Abcam, 18255) 1:2,000; anti-ubiquitinyl H2A antibody (Cell Signalling 8240) 1:2,000; anti-TET1 antibody (N terminus) (GeneTex GTX125888) 1:1,000; anti-lamin B antibody (C20) (Santa Cruz Biotechnologies, sc-6216) 1:10,000. Donkey anti-rabbit IgG-HRP (Santa Cruz Biotechnologies, sc-2077) or donkey anti-goat IgG-HRP (Santa Cruz Biotechnologies, sc-2056) secondary antibodies were then incubated for 1 h at room temperature. Blots were developed using Luminata Crescendo Western HRP substrate (EMD Millipore).

Statistics and reproducibility. No statistical methods were used to predetermine sample size, the experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment. All statistical tests are clearly described in the figure legends and/or in the Methods section, and exact *P* values values or adjusted *P* values are given where possible. For WGBS experiments (Fig. 3a, b, and Extended Data Figs 1a, 2c–e, 3a, b, 4, 7, 8), data are derived from cells from either *n* = 1 (E10.5 PGC sample) or *n* = 2 (all other samples) biological replicates and each replicate was obtained from pooled embryos (E10.5, *n* = 39 embryos from 4 litters; E11.5, *n* = 8 embryos from 1 litter; E12.5M/F, *n* = 4 embryos from 1 litter). For Aba-seq experiments (Figs 1d, 3a, b, and Extended Data Figs 1c–e, 2a–f, 3a, b, 4, 7, 8, 9b), data are derived from cells from *n* = 2 biological replicates and each replicate was obtained from pooled embryos (E10.5, *n* = 40 embryos from 4 litters; E11.5, *n* = 8 embryos from 1 litter; E12.5M/F, *n* = 4 embryos from 1 litter). For RNA-seq of mouse ES cells, samples are derived from *n* = 2 biological replicates corresponding to *n* = 2 independently cultured samples from one cell line. Complete details for PGC LC-MS, RNA-seq and RRBS data are reported in Supplementary Table 6 including the numbers of embryos and litters from which samples were derived. Western blots (Extended Data Figs 9f, 11b) were performed three times with similar results, and representative blots are shown. All immunostaining experiments (Figs 1e, 2a, b, Extended Data Fig. 6b) were performed twice with similar results and representative images are shown. Traditional bisulfite sequencing (Extended Data Fig. 6f, g) was carried out twice and a representative methylation profile is shown. For analysis of previously published WGBS (Extended Data Fig. 10a, b), TAB-seq (Extended Data Fig. 1c–e), Aba-seq (Extended Data Figs 1c–e, 2b, 10a, 10c) and ChIP-seq (Fig. 3b, Extended Data Figs 10a, 10c–g) datasets from mouse ES cells (see Methods for accession numbers), other than the H2Aub ChIP-seq dataset in which *n* = 1,

biological replicates were analysed both combined (shown) and separately (not shown) to ensure reproducibility of analysis.

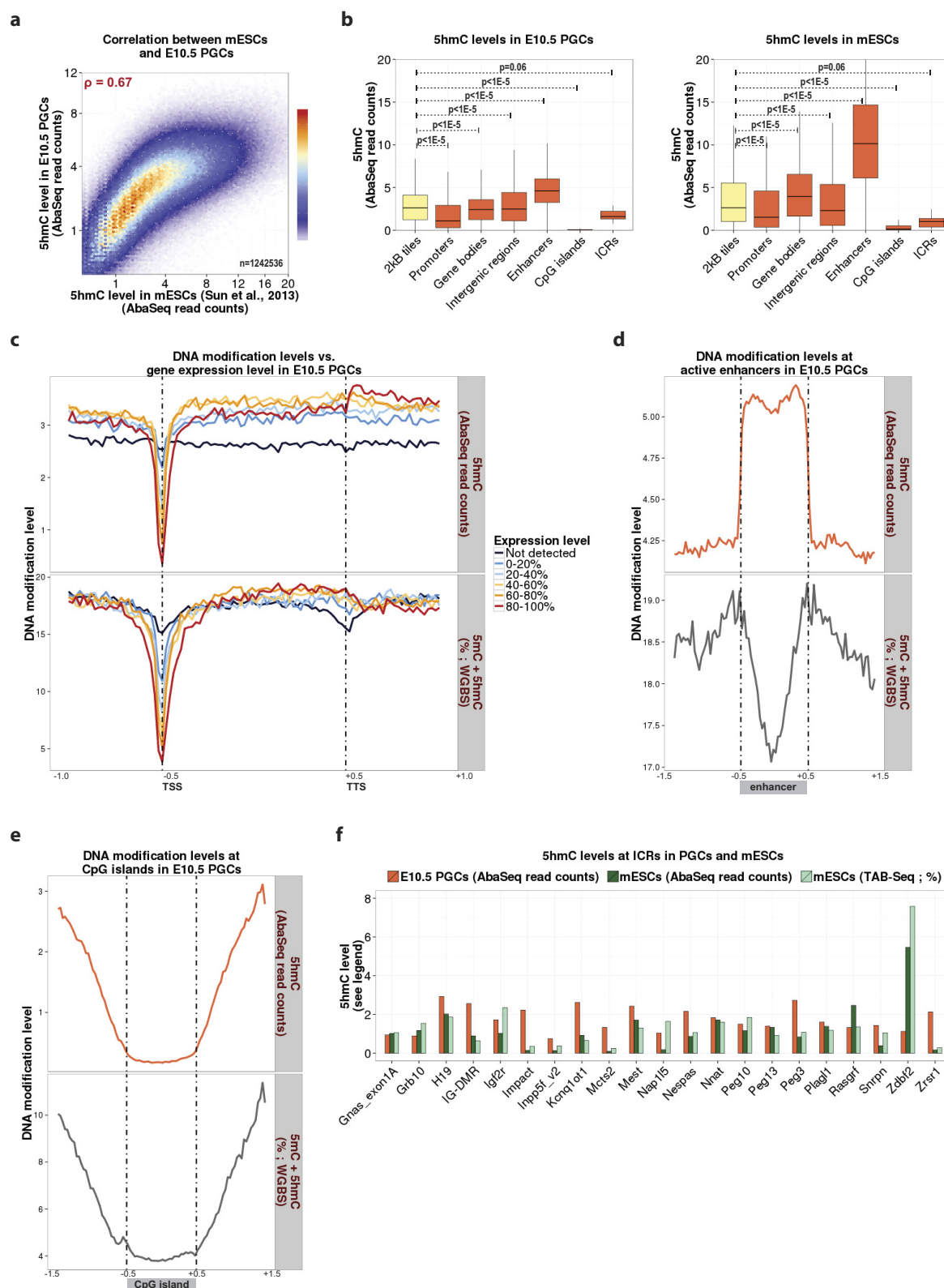
Data availability. Sequencing data reported in this paper are tabulated in Supplementary Tables 1–4 and are available at Gene Expression Omnibus (GEO) under accession GSE76973. All other data are available from the corresponding author upon reasonable request.

31. Yang, H. *et al.* One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell* **154**, 1370–1379 (2013).
32. Boyle, P. *et al.* Gel-free multiplexed reduced representation bisulfite sequencing for large-scale DNA methylation profiling. *Genome Biol.* **13**, R92 (2012).
33. Kurimoto, K. *et al.* Quantitative dynamics of chromatin remodeling during germ cell specification from mouse embryonic stem cells. *Cell Stem Cell* **16**, 517–532 (2015).
34. Borgel, J. *et al.* Targets and dynamics of promoter DNA methylation during early mouse development. *Nat. Genet.* **42**, 1093–1100 (2010).
35. Yu, M. *et al.* Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* **149**, 1368–1380 (2012).
36. Xu, Y. *et al.* Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Mol. Cell* **42**, 451–464 (2011).
37. Brookes, E. *et al.* Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs. *Cell Stem Cell* **10**, 157–170 (2012).
38. Cooper, S. *et al.* Targeting polycomb to pericentric heterochromatin in embryonic stem cells reveals a role for H2AK119u1 in PRC2 recruitment. *Cell Reports* **7**, 1456–1470 (2014).
39. Hashimoto, H. *et al.* Structure of a Naegleria Tet-like dioxygenase in complex with 5-methylcytosine DNA. *Nature* **506**, 391–395 (2014).



Extended Data Figure 1 | Characterization of WGBS datasets and validation of Aba-seq method. **a**, Distribution of WGBS coverage for each symmetric CpG. For box plots, the upper and lower hinges correspond to the first and third quartiles, the centre line corresponds to the median, and the maxima and minima correspond to the highest and lowest value within $1.5 \times$ the inter-quartile range, respectively. **b**, Overview of the Aba-seq method¹⁵. **c–e**, Density heat map showing correlation

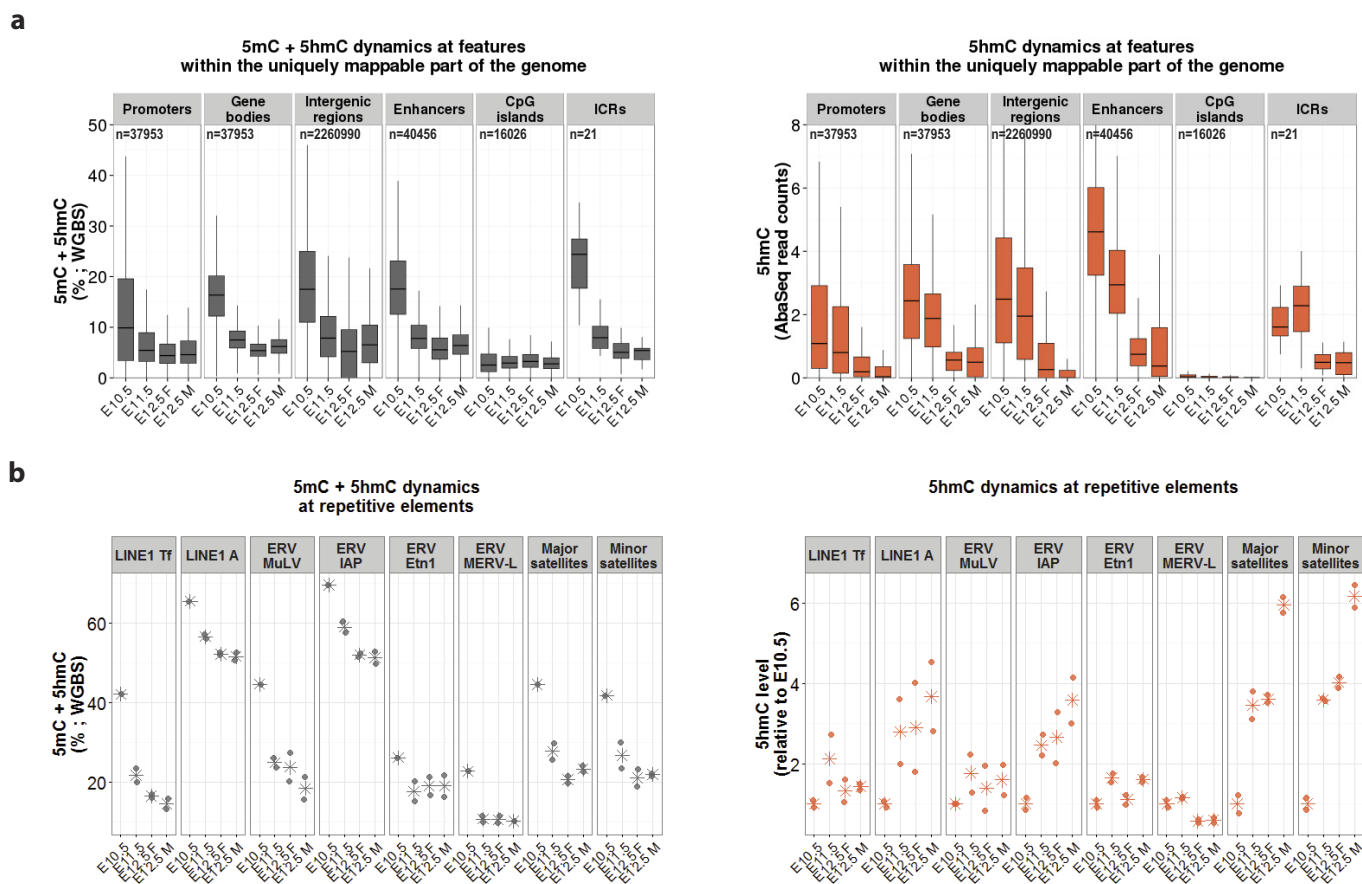
between levels of 5hmC at all 2-kb windows (minimum of four symmetric CpGs) in E14 mouse ES cells as computed by TAB-seq³⁵ (x-axis) and Aba-seq¹⁵ (y-axis) (**c**); TAB-seq³⁵ (x-axis) and hMeDIP³⁶ (y-axis) (**d**) or Aba-seq¹⁵ (x-axis) and hMeDIP³⁶ (y-axis) (**e**). For **c–e**, the Pearson correlation coefficient (ρ) is shown. Specific details for all Extended Data Figures regarding sample sizes and how samples were collected can be found in the ‘Statistics and reproducibility’ section.



Extended Data Figure 2 | Further analysis of the levels of 5hmC in E10.5 PGCs.

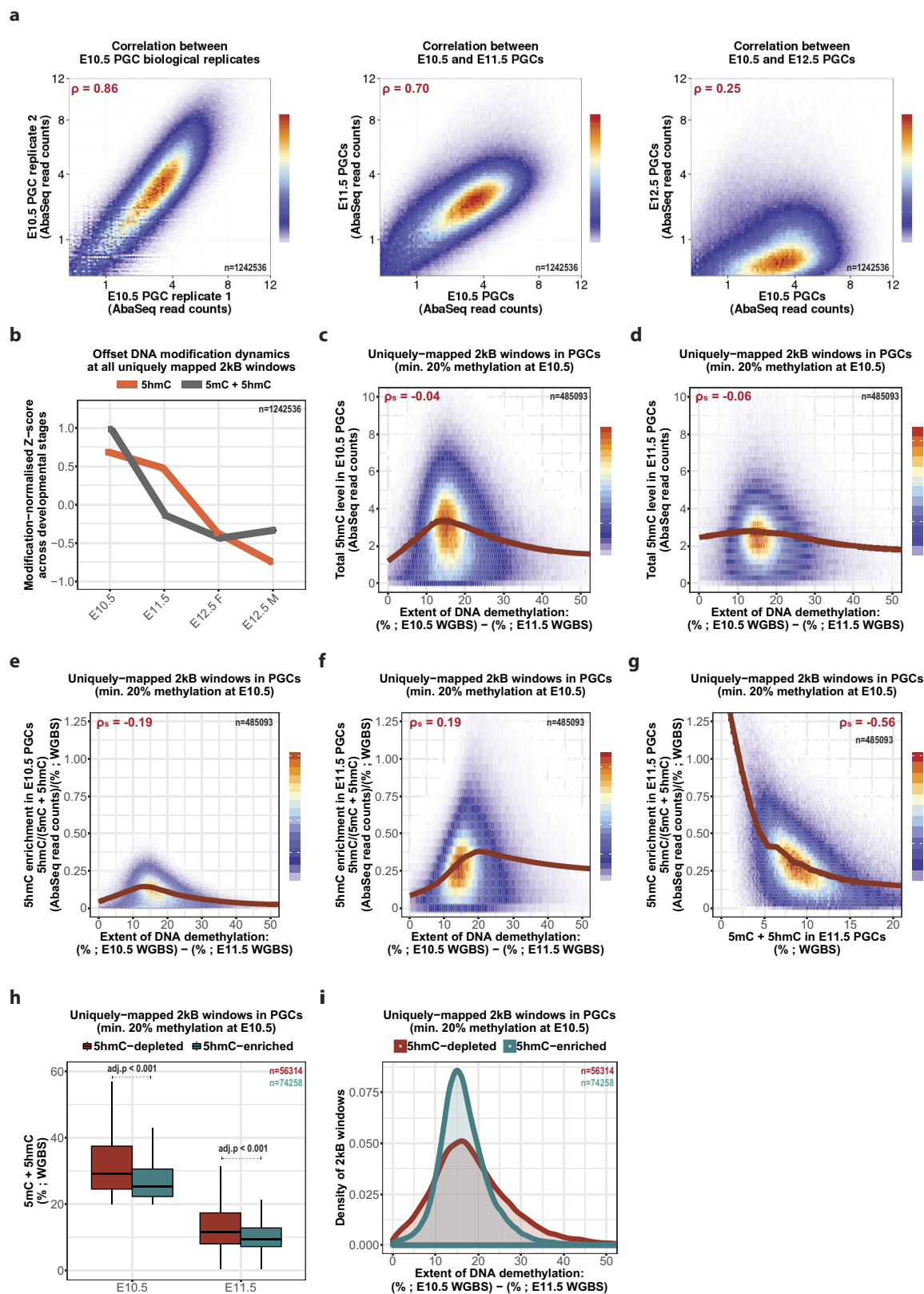
a, Density heat map showing the levels of 5hmC per 2-kb window (with minimum four CpGs) for E10.5 PGCs (y-axis) and E14 mouse ES cells¹⁵ (x-axis). The Pearson correlation coefficient (ρ) is shown. **b**, Levels of 5hmC (ascertained using Aba-seq) at various regulatory elements in E10.5 PGCs (left) or E14 mouse ES cells¹⁵ (right). P values are based on ANOVA and Dunnett's post hoc test. For box plots, the upper and lower hinges correspond to the first and third quartiles, the centre line corresponds to the median, and the maxima and minima correspond to the highest and lowest value within $1.5 \times$ the inter-quartile range,

respectively. **c**, Metagene plot showing the levels of 5hmC (determined using Aba-seq) (top) and the combined levels of 5mC and 5hmC (determined using WGBS) (bottom) in E10.5 PGCs across genes expressed at different levels in E10.5 PGCs. **d**, **e**, Metagene plot showing the levels of 5hmC (determined using Aba-seq) (top) and the combined levels of 5mC and 5hmC (determined using WGBS) (bottom) in E10.5 PGCs across either putative active enhancers (**d**) or CpG islands (**e**). **f**, Bar chart showing the levels of 5hmC at ICRs in E14 mouse ES cells as determined by TAB-seq³⁵ (%; light green) or Aba-seq¹⁵ (read counts; dark green), or in E10.5 PGCs as determined by Aba-seq (read counts; orange).



Extended Data Figure 3 | Further analysis of 5mC and 5hmC dynamics in PGCs. a, Combined levels of 5mC and 5hmC as determined by WGBS (left) or levels of 5hmC as determined by Aba-seq (right) at various features within the uniquely mapped part of the genome in PGCs between E10.5 and E12.5. For box plots, the upper and lower hinges correspond to the first and third quartiles, the centre line corresponds to the median, and

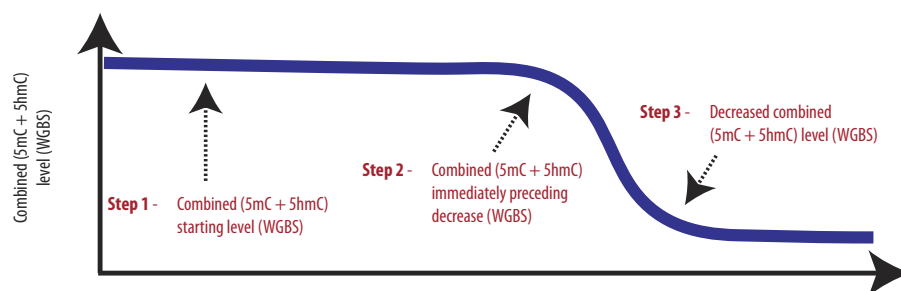
the maxima and minima correspond to the highest or lowest value within $1.5 \times$ the inter-quartile range, respectively. **b,** The combined levels of 5mC and 5hmC (determined by WGBS) (left) or levels of 5hmC (determined by Aba-seq) (right) at various consensus repetitive elements in PGCs between E10.5 and E12.5. Asterisks indicate mean values.



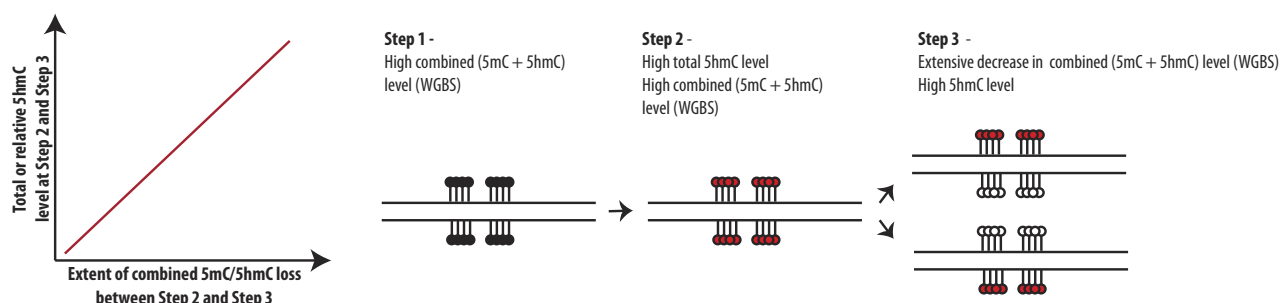
Extended Data Figure 4 | See next page for caption.

Extended Data Figure 4 | 5hmC is targeted to newly hypomethylated regions following DNA demethylation in mouse gonadal PGCs. See also Extended Data Fig. 5. **a**, Density heat map showing Pearson correlation (ρ) between levels of 5hmC for E10.5 PGC biological replicates (left), for E10.5 and E11.5 PGCs (middle), and for E10.5 and E12.5 PGCs (right). **b**, Mean Z-scores depicting levels of 5hmC (determined by Aba-seq) (orange) and combined levels of 5mC and 5hmC (determined by WGBS) (grey) for each stage normalized to the average level of either 5hmC or combined 5mC and 5hmC across stages. Standard error of the mean is shown but it is too small to see. **c–f**, Density heat maps showing the correlation between the total (**c**, **d**) or relative (**e**, **f**) levels of 5hmC in E10.5 (**c**, **e**) or E11.5 (**d**, **f**) PGCs and the change in the combined levels of 5mC and 5hmC in PGCs between these two stages for all 2-kb windows with a minimum 20% combined 5mC and 5hmC in E10.5 PGCs. **g**, Density heat map showing the correlation between the relative levels of 5hmC in E11.5 PGCs and the combined level of 5mC and 5hmC in E11.5 PGCs for all 2-kb windows with a minimum 20% combined 5mC and 5hmC in E10.5 PGCs. **h**, Combined levels of 5mC and 5hmC in E10.5 and E11.5 PGCs for 2-kb

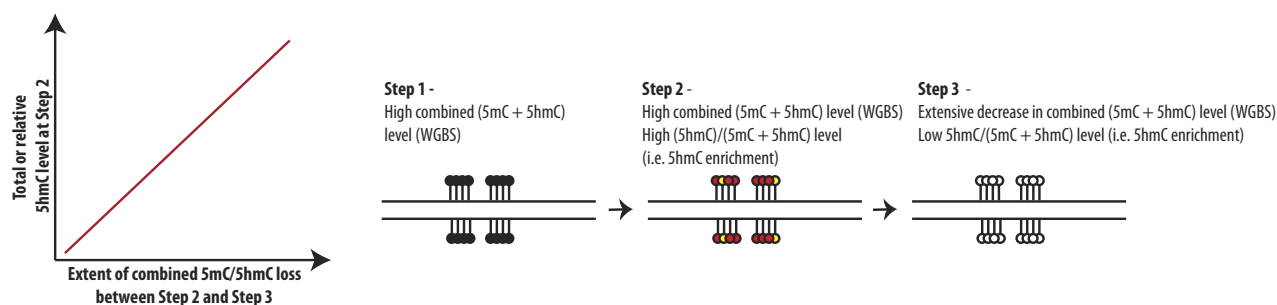
windows with a minimum 20% combined 5mC and 5hmC in E10.5 PGCs that are either 1) enriched for total levels of 5hmC at either E10.5 or E11.5 (green, upper-tail adjusted Poisson $P < 0.05$), or 2) depleted of total 5hmC at both E10.5 and E11.5 (red, lower-tail adjusted Poisson $P < 0.05$). **i**, Density plot showing the decrease in combined levels of 5mC and 5hmC in PGCs between E10.5 and E11.5 for 2-kb windows with a minimum 20% total DNA modification in E10.5 PGCs that are either 1) enriched for total levels of 5hmC at either E10.5 or E11.5 (green, upper-tail adjusted Poisson $P < 0.05$), or 2) depleted of total 5hmC at both E10.5 and E11.5 (red, lower-tail adjusted Poisson $P < 0.05$). For all box plots, the upper and lower hinges correspond to the first and third quartiles, the centre line corresponds to the median, and the maxima and minima correspond to the highest or lowest value within $1.5 \times$ the inter-quartile range, respectively. P values are based on a two-sided Wilcoxon test. Note that for density heat maps, the Spearman correlation (ρ_s) is shown and the red line represents the smoothed mean as determined by a generalized additive model.



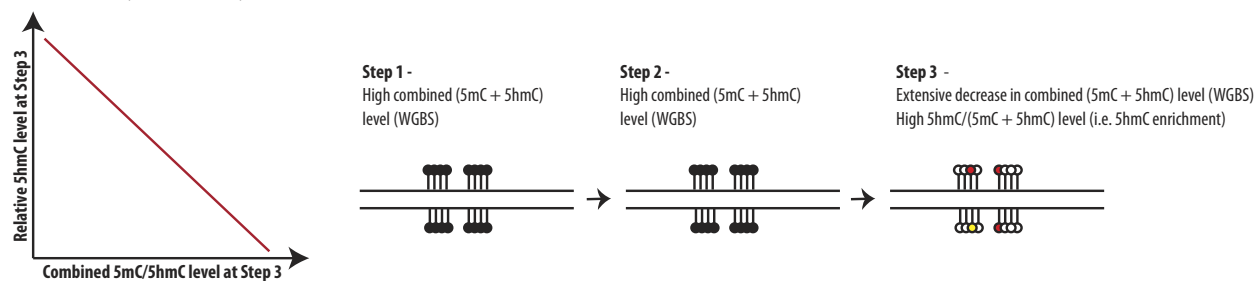
a Predictions for a model implicating 5hmC in passive DNA demethylation



b Predictions for a model implicating 5mC oxidation in initiation of active DNA demethylation



c Predictions for a model implicating 5mC oxidation in safeguarding the newly hypomethylated DNA state

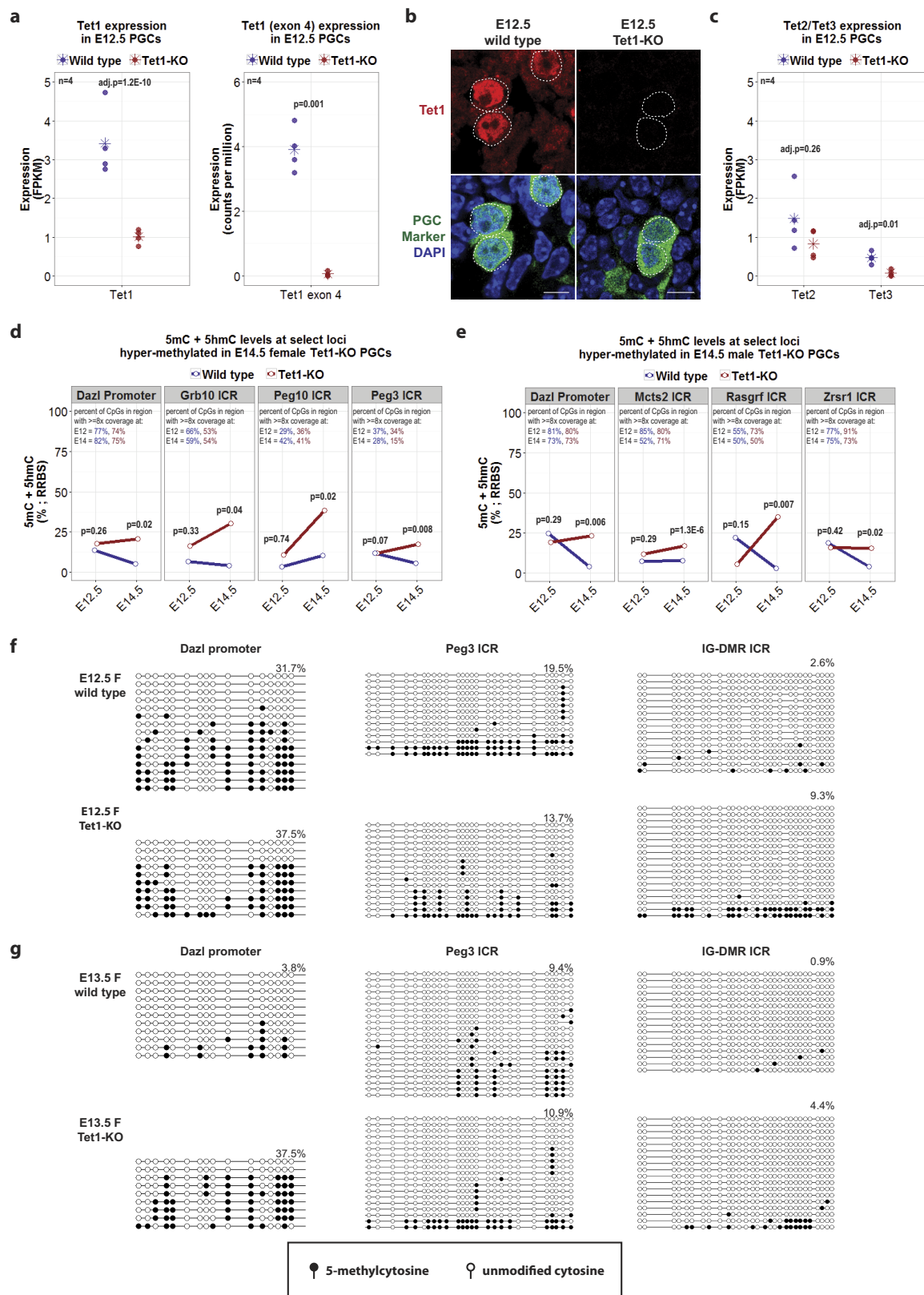


● 5-methylcytosine ○ unmodified cytosine ● 5-hydroxymethylcytosine ● 5-carboxycytosine

Extended Data Figure 5 | See next page for caption.

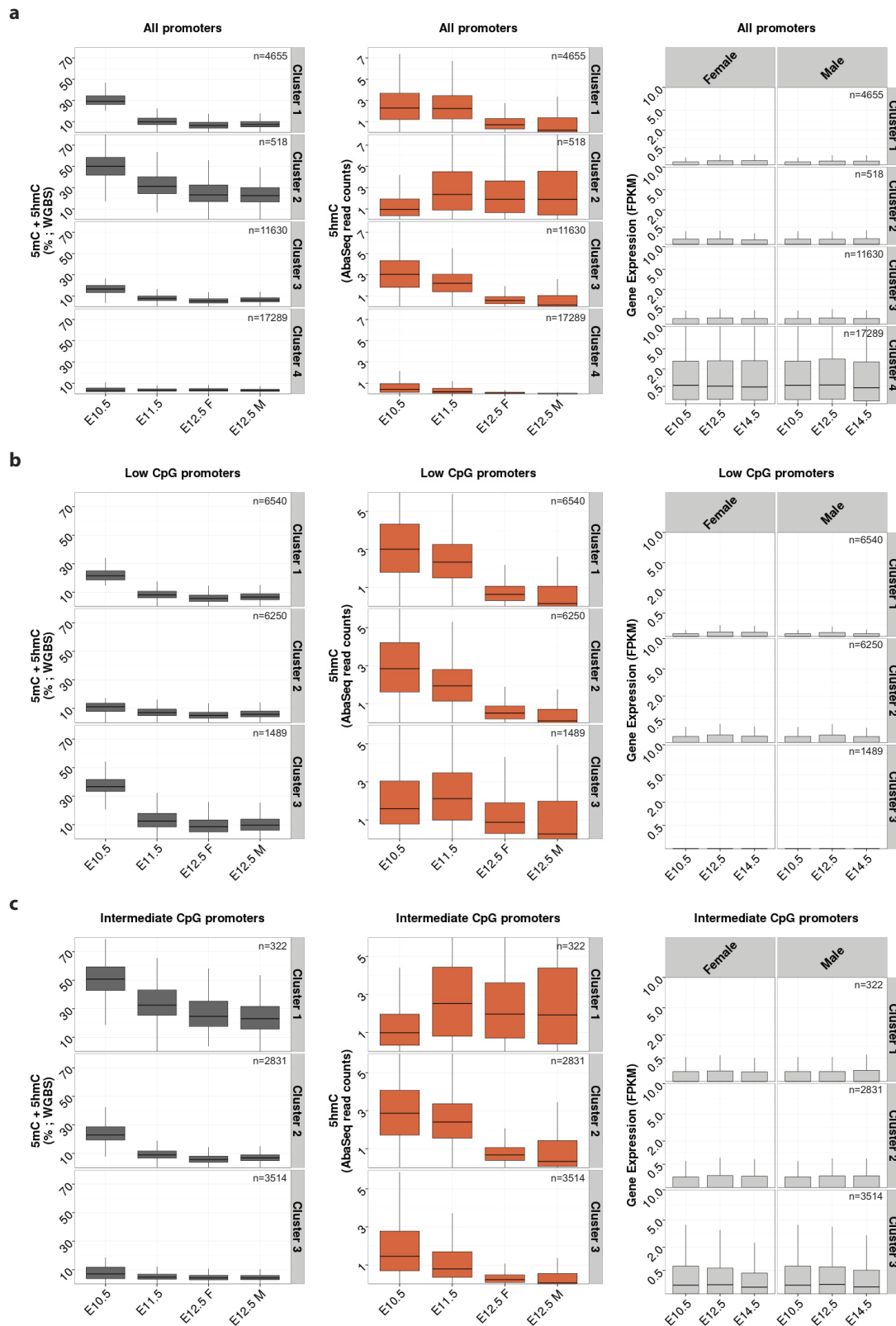
Extended Data Figure 5 | Suggested models implicating 5mC oxidation in DNA demethylation of gonadal PGCs. **a**, A model of oxidation followed by passive dilution predicts a positive correlation between the extent to which the combined levels of 5mC and 5hmC decrease between two stages (as determined by WGBS) and the total level of 5hmC at both the stage immediately preceding and following the decrease. **b**, A model implicating 5mC oxidation in triggering DNA demethylation via an active mechanism predicts a positive correlation between the extent to which the combined levels of 5mC and 5hmC decrease between two stages (as determined by WGBS) and the relative levels of 5hmC in the stage immediately preceding this decrease, as further oxidation of 5hmC to 5-formylcytosine (5fC) is the rate-limiting step in the full

oxidation of 5mC to 5-carboxylcytosine (5caC) (ref. 39). **c**, A model implicating oxidation of 5mC in safeguarding DNA hypomethylation following the major wave of DNA demethylation predicts that regions where the majority of DNA methylation has been lost between two stages (that is, those that are newly hypomethylated) will have high relative levels of 5hmC in the stage immediately after the major wave of DNA demethylation in order to remove residual methylation and/or aberrant *de novo* methylation. Thus, a limited correlation between the extent to which the combined levels of 5mC and 5hmC decrease between two stages (as determined by WGBS) and the relative levels of 5hmC in the stage immediately following this decrease may also be seen.



Extended Data Figure 6 | Expression of TET1, TET2 and TET3 and locus-specific DNA methylation in *Tet1*^{-/-} PGCs during epigenetic reprogramming. **a**, Expression of *Tet1* total transcript (left) or *Tet1* exon 4 (right) in E12.5 *Tet1*^{-/-} and wild-type PGCs. Adjusted P values (left) computed by DESeq2 and P values (right) computed by Student's t -test. Asterisks indicate mean values. **b**, Representative immunostaining against the N terminus of TET1 protein in E12.5 wild-type and *Tet1*^{-/-} PGCs. Scale bars represent 10 μ m. **c**, Expression of *Tet2* and *Tet3* in E12.5 *Tet1*^{-/-} and wild-type PGCs. Adjusted P values computed by DESeq2.

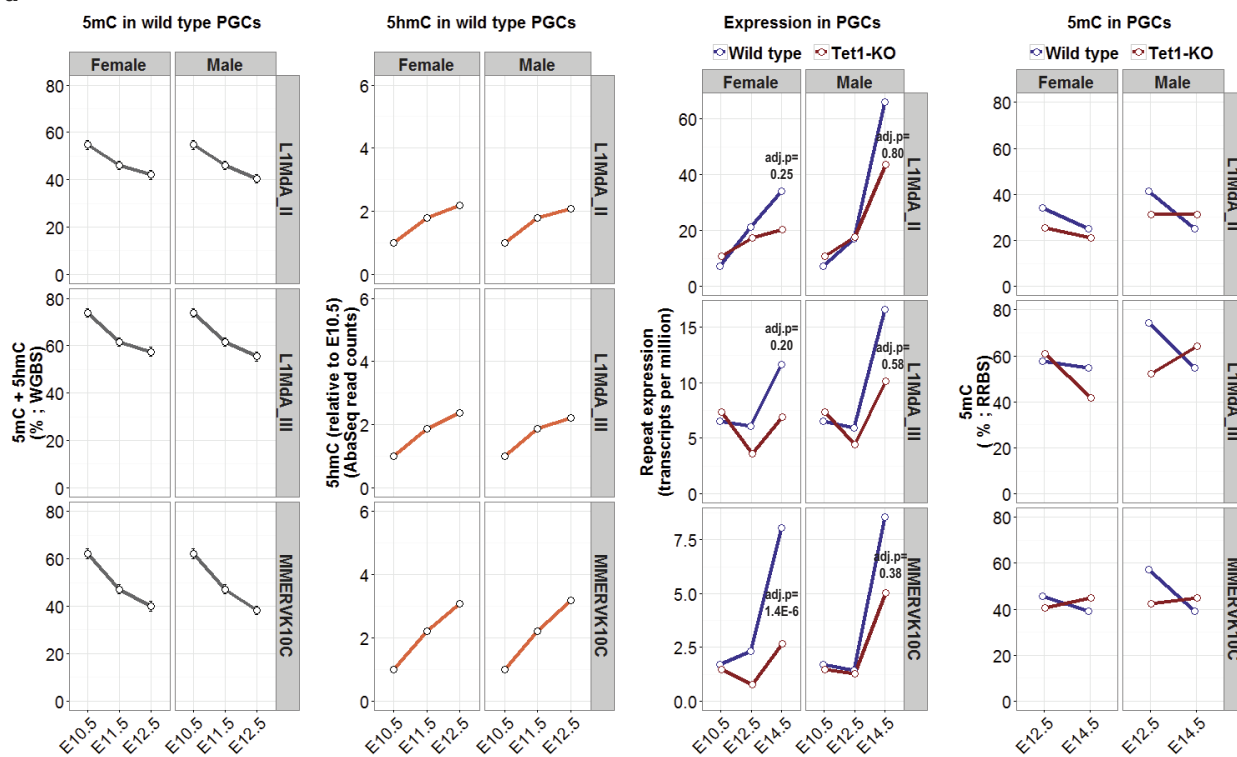
Asterisks indicate mean values. **d**, **e**, Mean combined levels of 5hmC and 5mC (determined using RRBS) in female (**d**) or male (**e**) E12.5 and E14.5 *Tet1*^{-/-} and wild-type PGCs for ICRs and germline gene promoters labelled as hypermethylated in E14.5 *Tet1*^{-/-} PGCs. The mean DNA modification level and P values were computed using RnBeads software (see Methods). **f**, **g**, Locus-specific bisulfite sequencing of the *Dazl* promoter (left), the *Peg3* ICR (middle) and the IG-DMR ICR (right) in E12.5 (**f**) and E13.5 (**g**) female *Tet1*^{-/-} and wild-type PGCs.



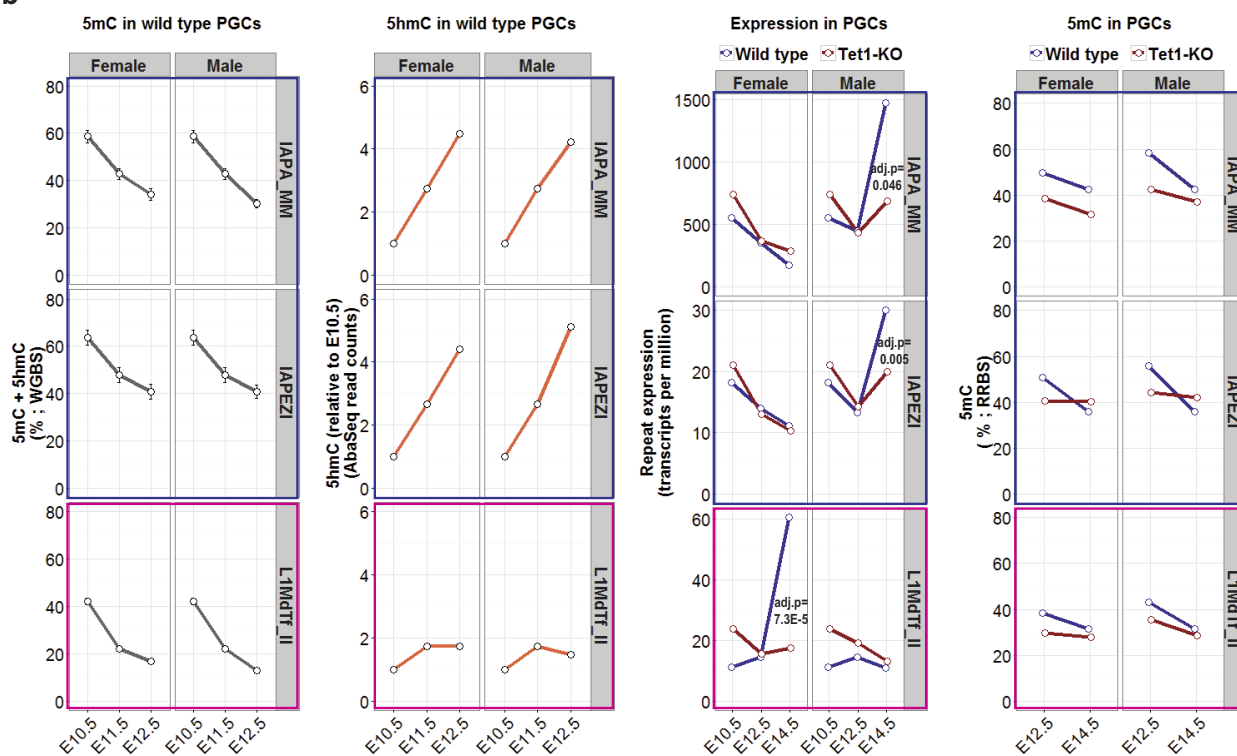
Extended Data Figure 7 | Promoter DNA methylation clustering analysis during germline reprogramming. **a**, The combined levels of 5mC and 5hmC at promoters (ascertained using WGBS) (left), levels of 5hmC at promoters (ascertained by Aba-seq) (centre), or gene expression levels (RNA-seq data) (right) in consecutive stages of PGC development for all genes grouped by *k*-means clustering of the combined 5mC and 5hmC dynamics at their promoter regions. **b**, **c**, Box plots depicting the combined levels of 5mC and 5hmC at promoters (ascertained using WGBS) (left), levels of 5hmC at promoters (ascertained using Aba-seq)

(centre), or gene expression levels (RNA-seq data) (right) in consecutive stages of PGC development for three clusters of genes with either low CpG promoters (**b**) or intermediate CpG promoters (**c**) grouped by *k*-means clustering of the combined 5mC and 5hmC dynamics at their promoter regions. For all box plots, the upper and lower hinges correspond to the first and third quartiles, the centre line corresponds to the median, and the maxima and minima correspond to the highest and lowest values within $1.5 \times$ the inter-quartile range, respectively.

a

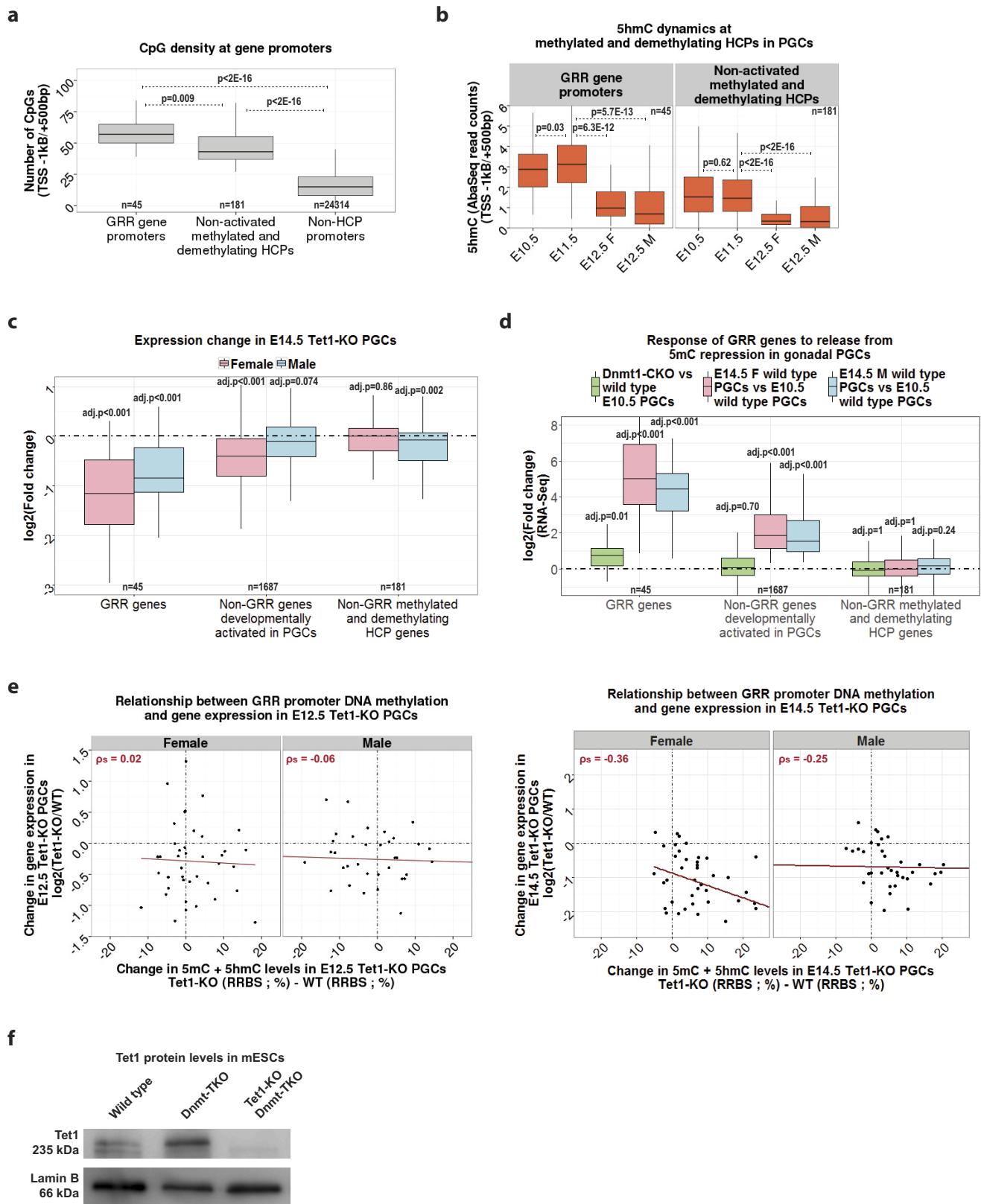


b



Extended Data Figure 8 | DNA modification and expression dynamics in wild-type and *Tet1*^{-/-} PGCs at retrotransposons normally activated concurrently with epigenetic reprogramming. a, b, Combined 5mC and 5hmC dynamics in wild-type PGCs (%; WGBS; far left); relative 5hmC dynamics (Aba-seq read counts normalized to E10.5) in wild-type PGCs (centre left); the expression dynamics in either wild-type or *Tet1*^{-/-} PGCs (transcripts per million (TPM); RNA-seq data; centre right); and the combined dynamics of 5mC and 5hmC in wild-type and *Tet1*^{-/-} PGCs

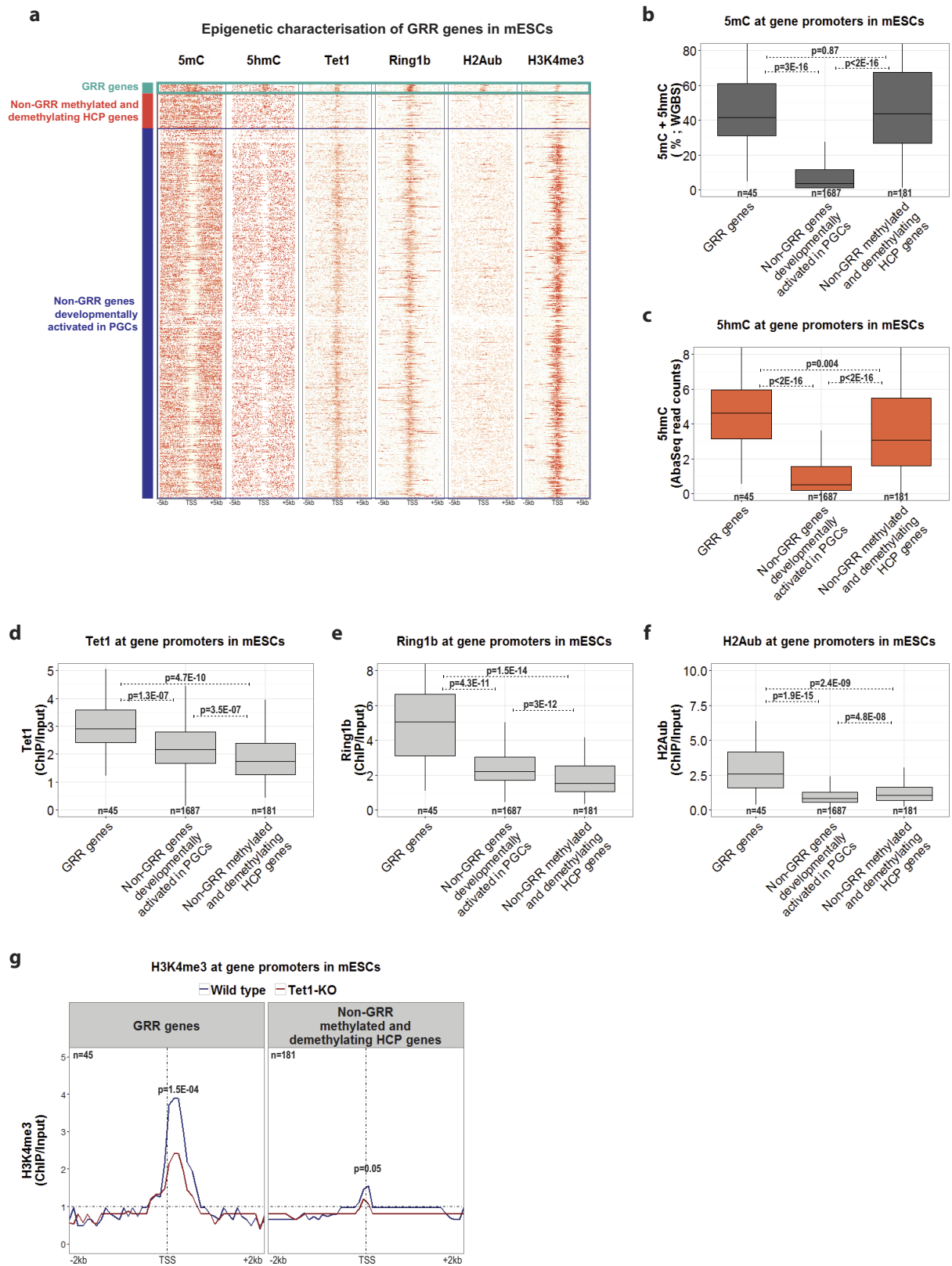
(%; RRBS; far right) for representative repetitive elements (IAPA_MM, IAPZ1 and L1MDTf_II) that are significantly upregulated (adjusted $P < 0.05$; Sleuth) in a sex-independent manner (a), a male-specific manner (b, blue outline) or a female-specific manner (b, pink outline) between E10.5 and E14.5 in wild-type PGCs. Mean values are shown in all cases. Adjusted P values for differential repeat expression analysis between E14.5 wild-type and *Tet1*^{-/-} PGCs are based on Sleuth software.



Extended Data Figure 9 | See next page for caption.

Extended Data Figure 9 | Characterization of GRR gene regulation by TET1 and 5mC in PGCs and mouse ES cells. **a**, CpG density at GRR gene promoters and other relevant promoters; *P* values are based on a two-sided Wilcoxon test. **b**, Mean 5hmC dynamics at GRR gene promoters and non-activated methylated and demethylated HCPs in PGCs; *P* values are based on a two-sided paired Wilcoxon test. **c**, \log_2 (fold change) between *Tet1*^{-/-} and wild-type E14.5 male (blue) or female (pink) PGCs for GRR genes and other relevant gene sets. FWER-adjusted *P* values are based on GSEA software (see Methods). **d**, \log_2 (fold change) between *Dnmt1*^{CKO} (ref. 24) and wild-type mouse PGCs (green), or between E14.5 female (pink) or male (blue) wild-type PGCs and E10.5 wild-type PGCs, for GRR genes and other relevant gene sets. FWER-adjusted *P* values are based on

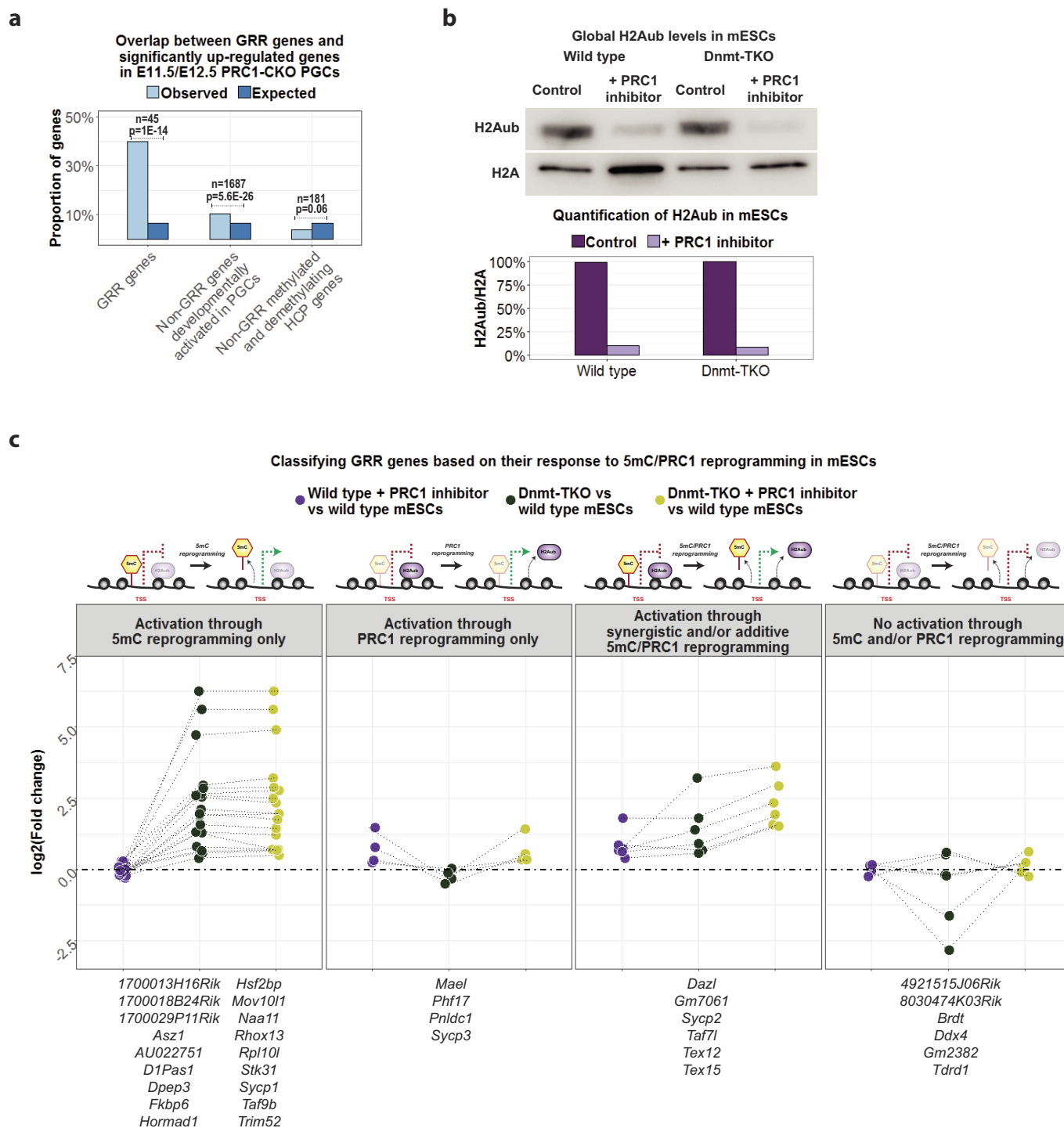
GSEA software (see Methods). **e**, Correlation between the difference in the combined levels of 5mC and 5hmC (%; ascertained by RRBS) (*x*-axis: *Tet1*^{-/-} – wild type (WT)) at GRR promoters and the change in GRR gene expression (*y*-axis; \log_2 (*Tet1*^{-/-}/wild type)) in E12.5 (left) and E14.5 (right) *Tet1*^{-/-} PGCs. Spearman's correlation is shown. **f**, Representative western blot showing TET1 and lamin B protein expression in wild-type, DNMT TKO, and *Tet1*^{-/-} DNMT TKO mouse ES cells. For all box plots, the upper and lower hinges correspond to the first and third quartiles, the centre line corresponds to the median, and the maxima and minima correspond to the highest or lowest value within 1.5 × the inter-quartile range, respectively.



Extended Data Figure 10 | See next page for caption.

Extended Data Figure 10 | Epigenetic characterization of GRR gene promoters in mouse ES cells. **a**, Genomic sequences centred on transcription start sites of GRR genes, non-GRR genes activated in both male and female PGCs between E10.5 and E14.5, and non-GRR methylated and demethylated HCP genes in wild-type mouse ES cells grown in serum-containing medium. Each horizontal line represents one gene; the intensity of red indicates the relative enrichment for the feature shown at the top of each column. The transcription start site and sequences 5 kb upstream and downstream of the transcription start site are shown. **b–f**, Box plots depicting the combined levels of 5mC and 5hmC (ascertained using WGBS)³⁰ (**b**); levels of 5hmC (ascertained using Aba-seq)¹⁵ (**c**); levels of TET1 (ChIP-seq data)²¹ (**d**); levels of RING1B (ChIP-seq data)³⁸ (**e**) and levels of H2Aub (ChIP-seq data)³⁷ (**f**) at the promoters

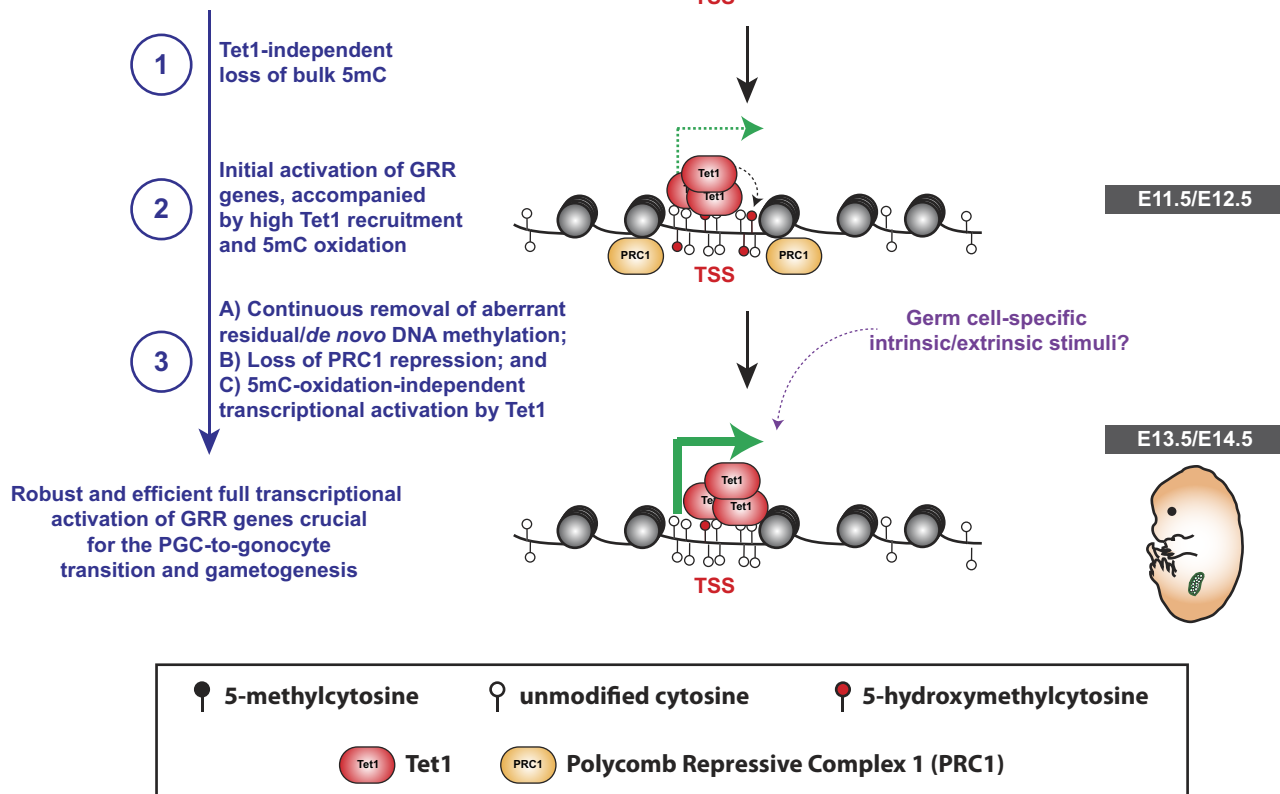
of GRR genes and of other relevant gene sets in wild-type mouse ES cells grown in serum-containing media. For all box plots, the upper and lower hinges correspond to the first and third quartiles, the centre line corresponds to the median, and the maxima and minima correspond to the highest and lowest value within $1.5 \times$ the inter-quartile range, respectively. *P* values are based on a two-sided Wilcoxon test. **g**, Metagene plot depicting median levels of H3K4me3 (ChIP-seq data)³⁰ around the transcription start sites of GRR genes (left) and non-GRR HCP genes that are also initially methylated and subsequently demethylated during PGC reprogramming (right) in wild-type and *Tet1*^{-/-} mouse ES cells grown in serum-containing medium. *P* values are based on a paired two-sided Wilcoxon test for the promoter (−1 kb/+500 bp) region.



Extended Data Figure 11 | Characterization of GRR gene regulation by PRC1 and 5mC in PGCs and mouse ES cells. **a**, Overlap between GRR genes and genes significantly upregulated in E11.5 and/or E12.5 PRC1 CKO PGCs compared with wild-type²⁶. *P* values based on hypergeometric test. **b**, Representative western blot showing H2Aub and H2A levels in

wild-type or DNMT TKO mouse ES cells after 6 h DMSO treatment, and wild-type or DNMT TKO mouse ES cells after 6 h PRT4165 treatment. **c**, Classification of GRR genes on the basis of their dependency for 5mC and/or PRC1 reprogramming in mouse ES cells (see Methods).

Transcriptional repression of germline reprogramming responsive (GRR) genes by 5mC and PRC1 (with simultaneous Tet1 recruitment and high 5hmC levels)



Extended Data Figure 12 | Model. The timely and efficient activation of GRR genes, involved in the transition from PGC to gonocyte and the correct progression of gametogenesis, requires interactions between promoter CpG density, the initiation of global DNA demethylation, TET1 recruitment, and removal of PRC1-mediated repression. Both DNA

demethylation-dependent (safeguarding against aberrant residual and/or *de novo* promoter DNA methylation) and -independent (such as the potential recruitment of OGT or other transcriptional activators to gene promoters^{22,28}) functions of TET1 are important for GRR gene activation.

Gating mechanisms of acid-sensing ion channels

Nate Yoder¹, Craig Yoshioka² & Eric Gouaux^{1,3}

Acid-sensing ion channels (ASICs) are trimeric¹, proton-gated^{2,3} and sodium-selective^{4,5} members of the epithelial sodium channel/degenerin (ENaC/DEG) superfamily of ion channels^{6,7} and are expressed throughout vertebrate central and peripheral nervous systems. Gating of ASICs occurs on a millisecond time scale⁸ and the mechanism involves three conformational states: high pH resting, low pH open and low pH desensitized⁹. Existing X-ray structures of ASIC1a describe the conformations of the open¹⁰ and desensitized¹¹ states, but the structure of the high pH resting state and detailed mechanisms of the activation and desensitization of the channel have remained elusive. Here we present structures of the high pH resting state of homotrimeric chicken (*Gallus gallus*) ASIC1a, determined by X-ray crystallography and single particle cryo-electron microscopy, and present a comprehensive molecular mechanism for proton-dependent gating in ASICs. In the resting state, the position of the thumb domain is further from the three-fold molecular axis, thereby expanding the 'acidic pocket' in comparison to the open and desensitized states. Activation therefore involves 'closure' of the thumb into the acidic pocket, expansion of the lower palm domain and an iris-like opening of the channel gate. Furthermore, we demonstrate how the $\beta 11$ – $\beta 12$ linkers that demarcate the upper and lower palm domains serve as a molecular 'clutch', and undergo a simple rearrangement to permit rapid desensitization.

To form well-ordered crystals of an ASIC in a high pH resting state, we used the $\Delta 25$ construct of chicken ASIC1a. This construct includes residues 25–463 of the full-length polypeptide and retains proton-dependent gating activity (Fig. 1a, b). Crystals of the $\Delta 25$ construct, belonging to the $P2_12_12_1$ space group, were grown at high pH in the presence of Ba^{2+} or Ca^{2+} cations¹² ($\Delta 25$ - Ba^{2+} and $\Delta 25$ - Ca^{2+} , respectively) and diffracted to 2.95 and 3.2 Å, respectively (Extended Data Table 1).

The $\Delta 25$ - Ba^{2+} and $\Delta 25$ - Ca^{2+} structures are nearly identical, and are consistent with the canonical chalice-like architecture of open¹⁰ and desensitized¹¹ channels. Individual subunits resemble a clenched fist composed of a transmembrane domain (TMD), palm, wrist, finger, knuckle, thumb and β -ball domains¹ (Fig. 1c). In comparison to open¹⁰ and desensitized¹¹ channels, the resting state structure positions the thumb and finger domains further from the three-fold molecular axis, resulting in an expanded extracellular domain (ECD) (Fig. 1d, e) and exposing an additional solvent-accessible surface area of approximately 595 Å² per subunit. The ion channel gate is closed and bears a notable resemblance to the TMD structure of the desensitized channel (Fig. 1e), indicating that the pore architecture is conserved across non-conducting functional states and suggesting that the conformation of the TMD is not directly pH-dependent.

Although the $\Delta 25$ channel is gated by protons, truncation of the amino and carboxy termini reduces its selectivity for Na^+ and decreases the Hill slope of proton-dependent ion channel activation (Extended Data Fig. 1). Point mutations on transmembrane helix 2b (TM2b), within the cytoplasmic region of the ion channel, also reduce sodium selectivity¹³. Given the altered function of the $\Delta 25$ construct and the importance of residues on or near the cytoplasmic domains of the ion

channel, we determined the structure of the full-length chicken ASIC1a channel in the resting state at high pH and in the presence of Ca^{2+} to a nominal resolution of 3.7 Å by cryo-electron microscopy (cryo-EM) (Fig. 2a–d, Extended Data Figs 2, 3, Extended Data Table 2).

Despite their functional differences, the $\Delta 25$ and full-length ASIC1a structures are almost identical at the present resolutions (Fig. 2e–g). Moreover, the superior quality of the cryo-EM density map in regions of the channel that exhibit low electron density in the X-ray structure provides valuable and unbiased structural information for domains that are important for gating and ion permeation in ASIC1a, including the acidic pocket and TMD (Fig. 2c, d). Nevertheless, key features of the X-ray structure are conserved in the cryo-EM structure, including an expanded acidic pocket and closed gate (Fig. 2c–g). Furthermore, in both the X-ray and cryo-EM structures, the TM2 helices undergo a domain swap that allows the Gly-Ala-Ser motif^{14–16} (also known as the GAS belt) to adopt an extended conformation immediately below the primary channel gate (Figs 1, 2, Extended Data Fig. 4). Thus, the conformation of the resting channel in our X-ray structures was not substantially affected by truncation of the cytoplasmic termini, model bias or crystal packing.

ASIC1a channels occupy a resting, closed state at physiological pH and activate within milliseconds in response to extracellular acidification^{17,18}. Consistent with its proposed role in proton-dependent gating^{1,19}, the acidic pocket, a solvent-exposed and electrostatically negative cavity formed at subunit interfaces, adopts an expanded conformation in both the X-ray and cryo-EM resting state structures. We hypothesize that this conformation is stabilized by hydrophobic and polar contacts across the finger, thumb and palm domains (Fig. 3a, b, Extended Data Fig. 5a, b). Upon extracellular acidification, thumb helices $\alpha 4$ and $\alpha 5$ shift towards the channel core as $\alpha 5$ undergoes a lateral pivot of 12° around its amino terminus, anchoring its carboxy terminus against the palm domain of a neighbouring subunit (Extended Data Fig. 5c). Rearrangements of thumb helices upon activation reduce the distance between titratable residues within the acidic pocket, enabling the formation of proton-mediated carboxyl–carboxylate pairings that stabilize the interface between the thumb, finger and palm domains.

The collapse of the acidic pocket upon exposure to protons is transduced to the channel pore via the palm domain, a network of β -strands that constitute the core of the channel, link movements of the ECD to the pore domain and frame extracellular fenestrations that provide access for cations to the extracellular vestibule and mouth of the pore (Extended Data Fig. 6). Activation initiates rearrangements across the ECD that manifest as counterclockwise rotations of individual subunits of approximately 5° around a lateral scaffold comprised of the β -ball and upper palm domains (Fig. 3c). The rotation of all subunits together leads to flexing of the lower palm towards the plasma membrane, displacing β -strand 1 ($\beta 1$) and $\beta 12$ in the palm domain by about 4 Å and inducing translation of TM1 and TM2a away from the three-fold molecular axis (Fig. 3d, Supplementary Video 1), culminating in the expansion of the extracellular fenestrations (Extended Data Fig. 6). The pore profile of the resting channel shows that it contains a closed

¹Vollum Institute, Oregon Health and Science University, Portland, Oregon 97239, USA. ²Department of Biomedical Engineering, Oregon Health and Science University, 2730 SW Moody Ave, Portland, Oregon 97201, USA. ³Howard Hughes Medical Institute, Oregon Health and Science University, Portland, Oregon 97239, USA.

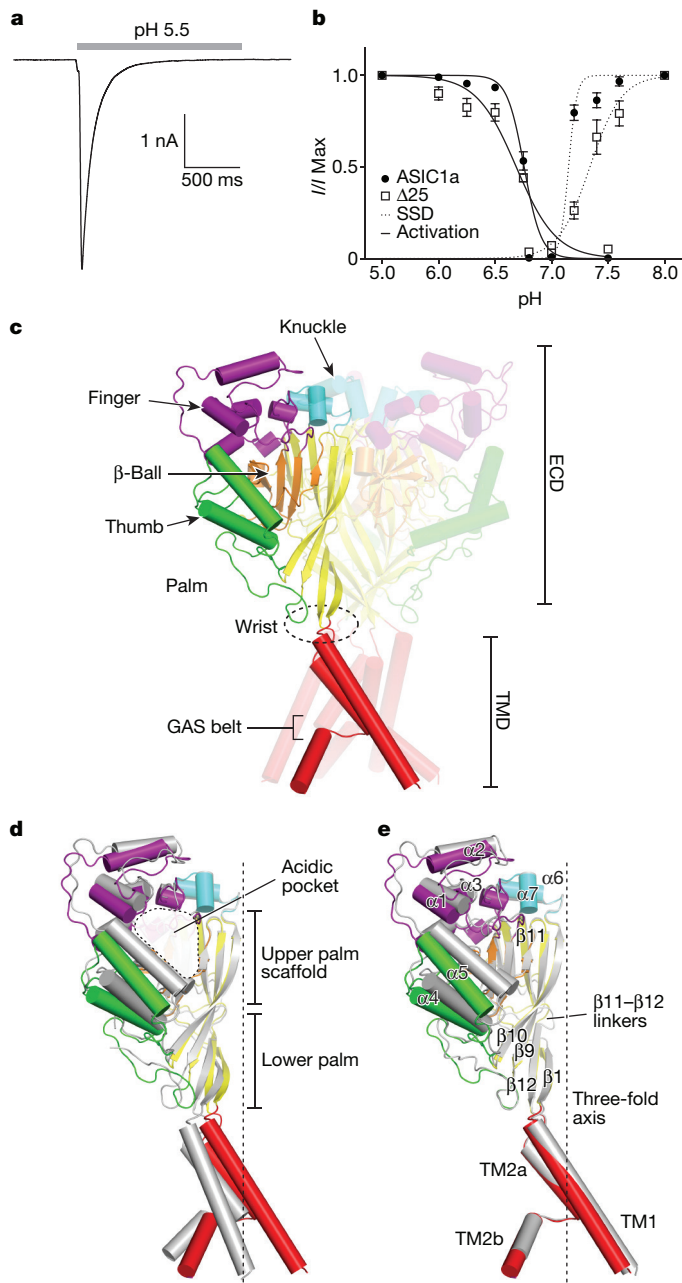


Figure 1 | Architecture and function of a resting channel.

a, b, Functional characterization of the $\Delta 25$ construct. Representative whole-cell patch-clamp recording (**a**) from $\Delta 25$ channels activated by step to low pH, and dose-response and steady-state desensitization (SSD) curves (**b**) for $\Delta 25$ and ASIC1a channels. Data were collected from Sf9 cells infected with BacMam virus containing $\Delta 25$ or ASIC1a DNA. Data are mean \pm s.e.m. Experiments were performed seven (activation) or five (SSD) times with similar results for ASIC1a, and ten (activation) or seven (SSD) times with similar results for $\Delta 25$. ASIC1a: $n = 7$ (activation) or 5 (SSD) cells; $\Delta 25$: $n = 10$ (activation) or 7 (SSD) cells. **c**, Structure of a $\Delta 25$ channel in the resting state at high pH, with different colours representing each domain. **d, e**, Single subunit superposition of resting channels with open (**d**; PDB code: 4NTW, grey) and desensitized (**e**; PDB code: 4NYK, grey) channels.

gate along the three-fold axis as a result of primary constrictions at Asp433 and Gly436 (Extended Data Fig. 7a, b). Proton-dependent rearrangements originating at the ECD facilitate channel activation via iris-like opening of transmembrane helices (Extended Data Fig. 7c, d), which shifts the carboxyl groups of Asp433 by 5.3 Å and opens the gate in the channel pore (Extended Data Fig. 7e, f).

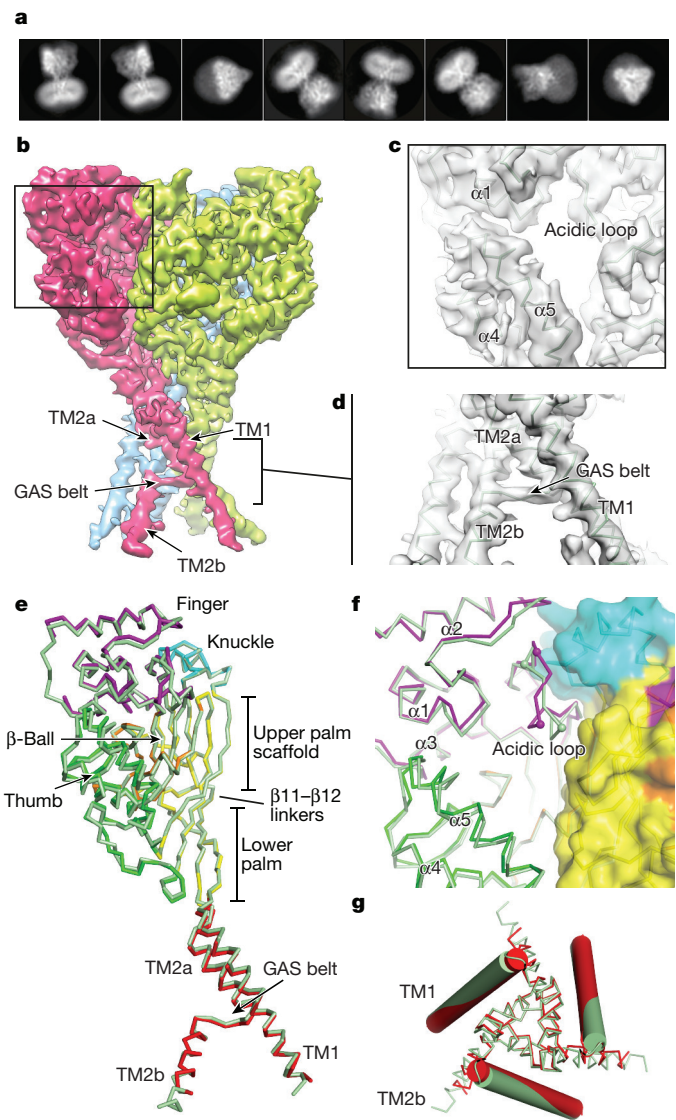


Figure 2 | Cryo-EM structure of chicken ASIC1a. **a–d**, Selected 2D classes (**a**) and cryo-EM density map (**b**) for ASIC1a with detail views of model and map of the acidic pocket (**c**) and the TM2 domain swap (**d**). **e**, Single subunit superposition of $\Delta 25$ and ASIC1a channels (light green). **f**, View of the acidic pocket from superposed $\Delta 25$ and ASIC1a channels (light green); Gly235 and Asp238 α -carbon atoms are offset by approximately 3 Å and are shown as spheres. **g**, Top-down view of the TMD from superposed $\Delta 25$ and ASIC1a channels (light green).

To elucidate the contributions of acidic pocket collapse to pH-dependent gating, we used site-directed double cysteine substitutions to introduce a disulfide bridge at residues Thr84 and Asn357 to anchor the thumb domain to the palm domain of a neighbouring subunit, thereby arresting the acidic pocket in an expanded conformation. In whole-cell patch-clamp experiments, reducing conditions recovered proton-dependent gating behaviour, increasing the magnitude of proton-dependent currents when compared to control conditions (Fig. 3e, f).

The observation that ASIC1a activation is blocked by an inter-subunit disulfide bond at the acidic pocket supports the hypothesized functional role of subunit–subunit interactions in ASIC1a channels²⁰, underscores the importance of the acidic pocket for ASIC1a gating and supports a simple gating scheme in which pH-dependent contraction of the acidic pocket drives channel activation.

ASIC1a channels undergo nearly complete desensitization on a timescale of hundreds of milliseconds^{17,18}. The $\beta 1$ – $\beta 2$ and $\beta 11$ – $\beta 12$

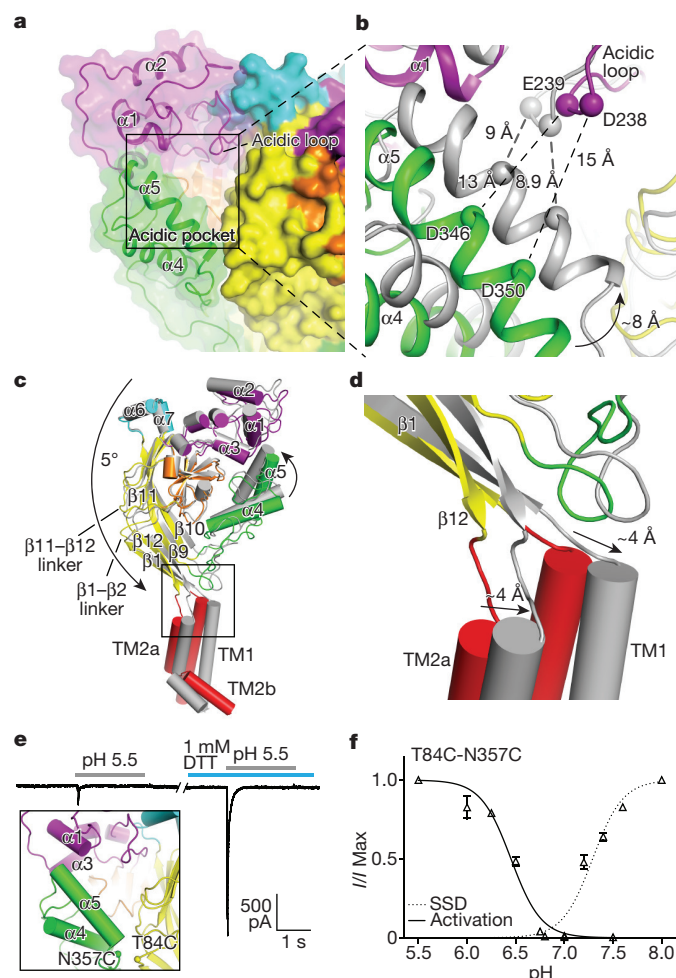


Figure 3 | Collapse of the acidic pocket initiates channel activation. **a**, **b**, Acidic pocket of the resting channel (**a**) and superposition with the open channel (**b**; PDB code: 4NTW, grey). **c**, **d**, Single subunit superposition of resting and open channels shows global (**c**) and lower-palm domain (**d**) conformational changes associated with channel activation. **e**, Whole-cell patch-clamp recording from T84C-N357C mutant channels. Data were recorded from CHO-K1 cells transfected with cDNA for the T84C-N357C mutant channel or Sf9 cells infected with BacMam virus containing DNA for the T84C-N357C mutant. Experiments were performed 11 times with similar results. Schematic representation of site-directed cysteine substitutions shown in inset. **f**, Dose-response and SSD curves (mean \pm s.e.m.) for T84C-N357C. Data were collected from Sf9 cells infected with BacMam virus containing T84C-N357C or ASIC1a DNA (ASIC1a is shown in Fig. 1b). For T84C-N357C channels, recordings were conducted in 1 mM dithiothreitol (DTT). ASIC1a: $n = 7$ (activation) or 5 (SSD) cells; T84C-N357C: $n = 7$ (activation) or 6 (SSD) cells. Experiments were performed seven (activation) or five (SSD) times with similar results for ASIC1a, and seven (activation) or six (SSD) times with similar results for $\Delta 25$.

linkers at the border of the upper and lower palm domains are important determinants of gating kinetics of ASIC channels^{21–25}. The overall conformation of the $\beta 1$ – $\beta 2$ and $\beta 11$ – $\beta 12$ linkers in the resting state mimics that in the open channel (Fig. 4a, b). In contrast to the similarities between resting and open states at $\beta 11$ – $\beta 12$ linkers, however, the side chains of Leu414 and Asn415 swap positions in the desensitized channel, resulting in a 9 Å reorientation of Leu414 towards the central vestibule and inducing a notable rearrangement of the $\beta 11$ – $\beta 12$ linkers²⁶ (Fig. 4c). We therefore propose that the conformation adopted by $\beta 11$ – $\beta 12$ linkers in the resting state provides a structural link between the upper and lower domains of the channel that enables pH-dependent collapse of the acidic pocket, which occurs approximately 40 Å from the plasma membrane, to drive activation.

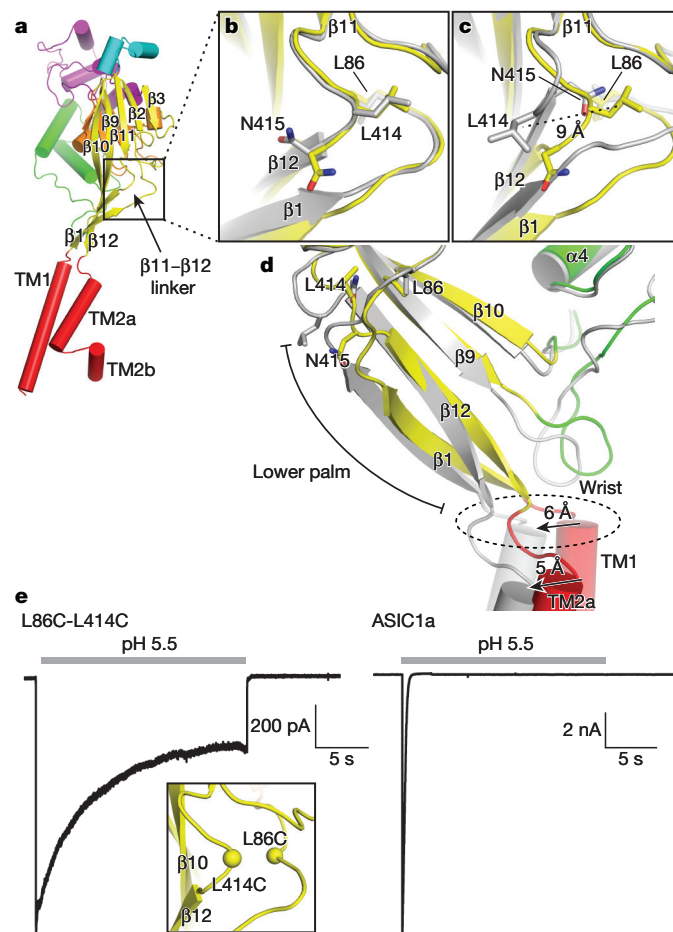


Figure 4 | A molecular clutch at $\beta 11$ – $\beta 12$ linkers controls desensitization. **a**, An individual subunit of an open ASIC1a channel (PDB code: 4NTW). **b**, **c**, Local alignments of an open channel with resting (**b**) and desensitized (**c**; PDB code: 4NYK) channels (both in grey) demonstrate conformational rearrangements associated with desensitization. **d**, Comparison of lower palm and TMDs for open and desensitized (grey) channels. **e**, Whole-cell patch-clamp recordings from L86C-L414C and ASIC1a channels. Inset shows a schematic of the site-directed cysteine substitutions. Data were collected from CHO-K1 cells transfected with cDNA for mutant or wild-type ASIC1a channels. Experiments were performed four times with similar results.

Furthermore, we hypothesize that rearrangement of the $\beta 11$ – $\beta 12$ linkers enables desensitization following prolonged exposure to low pH by serving as a molecular clutch, decoupling the collapsed acidic pocket from the lower channel and allowing TM1 and TM2a to relax by 6 Å and 5 Å, respectively (Supplementary Video 2, Fig. 4d), permitting the re-formation of the non-conducting ion channel. Accordingly, desensitization produces a ‘conformationally chimeric channel’ that bears a notable resemblance to the resting channel below, and to the open channel above, the $\beta 11$ – $\beta 12$ linkers (Fig. 1d). Upon return to physiological pH, expansion of the acidic pocket returns the $\beta 11$ – $\beta 12$ linkers to their original conformation, reforming a resting channel that is primed for subsequent activation.

To investigate the contribution of the rearrangement of the $\beta 11$ – $\beta 12$ linker to desensitization, we used site-directed double cysteine substitution to anchor Leu414 to an adjacent residue on $\beta 1$ – $\beta 2$ via a disulfide bridge. Exposure to protons elicited an inward current from L86C-L414C channels with an initial slow desensitizing component that gave way to a sustained current despite continued exposure to protons (Fig. 4e). Taken together, these data are consistent with Leu414 and the $\beta 11$ – $\beta 12$ linkers separating from the neighbouring $\beta 1$ – $\beta 2$ linkers upon continued exposure to protons, and with this region having a central role in the mechanism of desensitization.

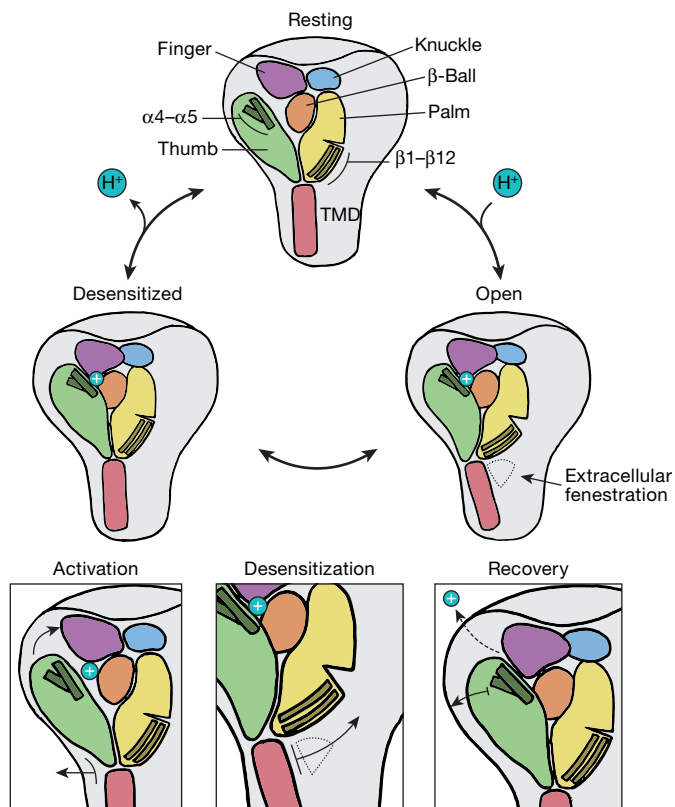


Figure 5 | Gating scheme. Cartoon representation of the proton-dependent gating cycle in ASIC1a channels.

Here we present X-ray and cryo-EM structures of ASIC1a channels in a resting state at high pH and in the presence of inhibitory Ba^{2+} or Ca^{2+} cations. The structure of a resting channel, the last remaining unsolved functional state of ASIC1a, consolidates our molecular understanding of the canonical functional states of ASIC1a channels and informs a comprehensive molecular model of pH-dependent gating mechanics (Fig. 5). Moreover, our data suggest that the TM2 domain swap and GAS belt are functional characteristics of ASIC1a architecture.

At physiological pH, ASIC1a channels predominantly occupy a resting state with a pore that is closed to ion permeation and an expanded acidic pocket. Upon exposure to low pH, the acidic pocket adopts a collapsed conformation as $\alpha 5$ pivots towards the channel core to enable proton-mediated carboxyl–carboxylate pairings between thumb and finger domains. The collapse of the acidic pocket initiates coordinated movements throughout the ECD that manifest as lateral rotations of individual subunits around their upper palm domain scaffold. Simultaneous rotation of all subunits displaces the $\beta 1$ and $\beta 12$ strands of the lower palm domain towards the membrane and away from the three-fold molecular axis. This shift of the lower palm domain results in expansion of the extracellular fenestrations and an iris-like opening of the channel gate, enabling ions to pass through the channel pore.

ASIC1a channels undergo rapid and complete desensitization at low pH. Continued exposure to protons results in a swap in sidechain orientations of Leu414 and Asn415 residues, inducing a substantial rearrangement of the $\beta 11$ – $\beta 12$ linkers that demarcate the upper and lower palm domains. Reorganization of palm domain linkers uncouples the low pH conformation of the upper ECD from the lower part of the channel, allowing transmembrane helices to relax back into a resting-like conformation and forming a desensitized channel that is insensitive to protons. Re-priming ASIC1a channels for activation requires the removal of protons. We hypothesize that, upon return to physiological pH values, electrostatic repulsion stemming from the

deprotonation of titratable acidic residues drives the expansion of the acidic pocket, enabling $\beta 11$ – $\beta 12$ to revert back to a non-swapped conformation and recovering proton sensitivity upon formation of a resting channel.

Our results build on previous studies of gating mechanisms in ASICs that suggest that activation induces movements at the acidic pocket and that the lower palm domain and $\beta 11$ – $\beta 12$ linkers have roles in desensitization^{27–29}. By contrast, atomic force microscopy studies³⁰ have indicated that the ECD of human ASIC1a channels increases in height upon activation, a conformational change that was not seen in our analysis of X-ray crystallographic and cryo-EM structures of chicken ASIC1a. Nevertheless, the high cooperativity with which protons activate ASIC1a is consistent with multiple resting states that are characterized by varying degrees of protonation, suggesting that these structures represent an average over a range of protein conformations. As ASICs are the best structurally characterized members of the epithelial sodium channel/degenerin superfamily of ion channels, our studies provide fundamental insights into the mechanisms of gating and modulation of the entire superfamily.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 July 2017; accepted 15 January 2018.

Published online 7 March 2018.

- Jasti, J., Furukawa, H., Gonzales, E. B. & Gouaux, E. Structure of acid-sensing ion channel 1 at 1.9 Å resolution and low pH. *Nature* **449**, 316–323 (2007).
- Krishtal, O. A. & Pidoplichko, V. I. A receptor for protons in the nerve cell membrane. *Neuroscience* **5**, 2325–2327 (1980).
- Waldmann, R., Champigny, G., Bassilana, F., Heurteaux, C. & Lazdunski, M. A proton-gated cation channel involved in acid-sensing. *Nature* **386**, 173–177 (1997).
- Bässler, E. L., Ngo-Anh, T. J., Geisler, H. S., Ruppersberg, J. P. & Gründer, S. Molecular and functional characterization of acid-sensing ion channel (ASIC) 1b. *J. Biol. Chem.* **276**, 33782–33787 (2001).
- Yermolaieva, O., Leonard, A. S., Schnitzler, M. K., Abboud, F. M. & Welsh, M. J. Extracellular acidosis increases neuronal cell calcium by activating acid-sensing ion channel 1a. *Proc. Natl Acad. Sci. USA* **101**, 6752–6757 (2004).
- Kellenberger, S. & Schild, L. Epithelial sodium channel/degenerin family of ion channels: a variety of functions for a shared structure. *Physiol. Rev.* **82**, 735–767 (2002).
- Waldmann, R. & Lazdunski, M. H^+ -gated cation channels: neuronal acid sensors in the NaC/DEG family of ion channels. *Curr. Opin. Neurobiol.* **8**, 418–424 (1998).
- Hesselager, M., Timmermann, D. B. & Ahning, P. K. pH dependency and desensitization kinetics of heterologously expressed combinations of acid-sensing ion channel subunits. *J. Biol. Chem.* **279**, 11006–11015 (2004).
- Zhang, P., Sigworth, F. J. & Canessa, C. M. Gating of acid-sensitive ion channel-1: release of Ca^{2+} block vs. allosteric mechanism. *J. Gen. Physiol.* **127**, 109–117 (2006).
- Baconguis, I., Böhlen, C. J., Goehring, A., Julius, D. & Gouaux, E. X-ray structure of acid-sensing ion channel 1–snake toxin complex reveals open state of a Na^+ -selective channel. *Cell* **156**, 717–729 (2014).
- Gonzales, E. B., Kawate, T. & Gouaux, E. Pore architecture and ion sites in acid-sensing ion channels and P2X receptors. *Nature* **460**, 599–604 (2009).
- Babini, E., Paukert, M., Geisler, H. S. & Gründer, S. Alternative splicing and interaction with di- and polyvalent cations control the dynamic range of acid-sensing ion channel 1 (ASIC1). *J. Biol. Chem.* **277**, 41597–41603 (2002).
- Lynagh, T. et al. A selectivity filter at the intracellular end of the acid-sensing ion channel pore. *eLife* **6**, e24630 (2017).
- Kellenberger, S., Hoffmann-Pochon, N., Gautschi, I., Schneeberger, E. & Schild, L. On the molecular basis of ion permeation in the epithelial Na^+ channel. *J. Gen. Physiol.* **114**, 13–30 (1999).
- Kellenberger, S., Gautschi, I. & Schild, L. A single point mutation in the pore region of the epithelial Na^+ channel changes ion selectivity by modifying molecular sieving. *Proc. Natl Acad. Sci. USA* **96**, 4170–4175 (1999).
- Snyder, P. M., Olson, D. R. & Bucher, D. B. A pore segment in DEG/ENaC Na^+ channels. *J. Biol. Chem.* **274**, 28484–28490 (1999).
- Zhang, P. & Canessa, C. M. Single channel properties of rat acid-sensitive ion channel-1 α , -2 α , and -3 expressed in *Xenopus* oocytes. *J. Gen. Physiol.* **120**, 553–566 (2002).
- Sutherland, S. P., Benson, C. J., Adelman, J. P. & McCleskey, E. W. Acid-sensing ion channel 3 matches the acid-gated current in cardiac ischemia-sensing neurons. *Proc. Natl Acad. Sci. USA* **98**, 711–716 (2001).
- Sherwood, T. et al. Identification of protein domains that control proton and calcium sensitivity of ASIC1a. *J. Biol. Chem.* **284**, 27899–27907 (2009).

20. Gwiazda, K., Bonifacio, G., Vullo, S. & Kellenberger, S. Extracellular subunit interactions control transitions between functional states of acid-sensing ion channel 1a. *J. Biol. Chem.* **290**, 17956–17966 (2015).
21. Li, T., Yang, Y. & Canessa, C. M. Leu85 in the $\beta 1$ – $\beta 2$ linker of ASIC1 slows activation and decreases the apparent proton affinity by stabilizing a closed conformation. *J. Biol. Chem.* **285**, 22706–22712 (2010).
22. Li, T., Yang, Y. & Canessa, C. M. Asn415 in the $\beta 11$ – $\beta 22$ linker decreases proton-dependent desensitization of ASIC1. *J. Biol. Chem.* **285**, 31285–31291 (2010).
23. Li, T., Yang, Y. & Canessa, C. M. Two residues in the extracellular domain convert a nonfunctional ASIC1 into a proton-activated channel. *Am. J. Physiol. Cell Physiol.* **299**, C66–C73 (2010).
24. Coric, T., Zhang, P., Todorovic, N. & Canessa, C. M. The extracellular domain determines the kinetics of desensitization in acid-sensitive ion channel 1. *J. Biol. Chem.* **278**, 45240–45247 (2003).
25. Springauf, A., Bresenitz, P. & Gründer, S. The interaction between two extracellular linker regions controls sustained opening of acid-sensing ion channel 1. *J. Biol. Chem.* **286**, 24374–24384 (2011).
26. Bacongus, I. & Gouaux, E. Structural plasticity and dynamic selectivity of acid-sensing ion channel–spider toxin complexes. *Nature* **489**, 400–405 (2012).
27. Vullo, S. *et al.* Conformational dynamics and role of the acidic pocket in ASIC pH-dependent gating. *Proc. Natl Acad. Sci. USA* **114**, 3768–3773 (2017).
28. Bonifacio, G., Lelli, C. I. & Kellenberger, S. Protonation controls ASIC1a activity via coordinated movements in multiple domains. *J. Gen. Physiol.* **143**, 105–118 (2014).
29. Ramaswamy, S. S., MacLean, D. M., Gorfe, A. A. & Jayaraman, V. Proton-mediated conformational changes in an acid-sensing ion channel. *J. Biol. Chem.* **288**, 35896–35903 (2013).
30. Yokokawa, M., Carnally, S. M., Henderson, R. M., Takeyasu, K. & Edwardson, J. M. Acid-sensing ion channel (ASIC) 1a undergoes a height transition in response to acidification. *FEBS Lett.* **584**, 3107–3110 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Goehring, D. Claxton and I. Bacongus for initial construct screening and advice through all aspects of the project, L. Vaskalis for help with figures, H. Owen for manuscript preparation, all members of the Gouaux laboratory for their support and the Berkeley Center for Structural Biology and the Northeastern Collaborative Access Team for help with X-ray data collection. This research was supported by the National Institute of General Medical Sciences (5T32DK007680), and the National Institute of Neurological Disorders and Stroke (5F31NS096782 to N.Y. and 5R01NS038631 to E.G.). Additional support was provided by ARCS Foundation and Tartar Trust fellowships. E.G. is an Investigator with the Howard Hughes Medical Institute.

Author Contributions N.Y. and E.G. designed the project. N.Y. performed biochemistry, crystallography and electrophysiology experiments. N.Y. and C.Y. performed cryo-EM data collection. C.Y. performed the cryo-EM data analysis. N.Y. wrote the manuscript and all authors edited the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to E.G. (gouauxe@ohsu.edu).

Reviewer Information *Nature* thanks L. Rash and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Receptor construct, expression and purification. The $\Delta 25$ crystallization construct has 24 residues removed from the amino terminus and 64 residues removed from the carboxy terminus. Recombinant $\Delta 25$ protein was expressed in HEK293S GnTI[−] cells by way of baculovirus-mediated gene transduction. HEK293S GnTI[−] cells in suspension culture were grown to a density of $3.5 \times 10^6 \text{ ml}^{-1}$ and infected with P3 virus. After 8 h of culture at 37 °C, sodium butyrate was added to 10 mM final concentration and the temperature was decreased to 30 °C. After 48 h of expression, cells were collected by centrifugation, washed with phosphate buffered saline (PBS), and resuspended in Tris-buffered saline containing protease inhibitors (TBS; 150 mM NaCl, 20 mM Tris pH 8.0, 1 mM phenylmethylsulfonyl fluoride, 0.05 mg ml^{−1} aprotinin, 2 $\mu\text{g ml}^{-1}$ pepstatin A and 2 $\mu\text{g ml}^{-1}$ leupeptin). Cells were disrupted by sonication and membrane fractions were isolated by ultracentrifugation. Full-length chicken ASIC1a channels were expressed in cell culture in the same way as the $\Delta 25$ construct but without the membrane fraction isolation step.

Membrane pellets containing $\Delta 25$ were resuspended in TBS buffer with protease inhibitors, homogenized and solubilized in 40 mM *n*-dodecyl β -D-maltoside (DDM) for 1 h at 4 °C. ASIC1a channels were solubilized in an identical manner immediately after cell disruption by sonication. The solubilized material was clarified by ultracentrifugation and the supernatant was incubated in metal ion affinity resin for 1.5 h at 4 °C with 10 mM imidazole. Co²⁺ resin was packed into a column and subjected to three column volume wash steps with buffer containing 300 mM NaCl, 20 mM Tris pH 8.0, 1 mM DDM and 10 mM imidazole followed by three additional column volume washes with buffer containing 300 mM NaCl, 20 mM Tris pH 8.0, 1 mM DDM and 30 mM imidazole. Bound protein was eluted with 300 mM NaCl, 20 mM Tris pH 8.0, 1 mM DDM and 250 mM imidazole. The histidine-tagged enhanced green fluorescent protein (eGFP) tag was cleaved by thrombin digestion. The $\Delta 25$ protein was further purified by size-exclusion chromatography (SEC) using a mobile phase containing 300 mM NaCl, 20 mM Tris pH 8.0, 2 mM *n*-decyl β -D-thiomaltopyranoside (C10ThioM), 1 mM DTT, 0.2 mM cholesteryl hemisuccinate (CHS) and 5 mM CaCl₂ or BaCl₂. Peak fractions were collected and concentrated to $\sim 3 \text{ mg ml}^{-1}$.

Crystallization. Cholesterol stocks were maintained at 50 mg ml^{−1} in chloroform and stored at −20 °C. Cholesterol aliquots were removed from stock and placed under argon until visibly dry. Dried cholesterol aliquots of 6 mg were resuspended by adding 20 μl of an aqueous solution of 400 mM C10ThioM, followed by gentle stirring for 1 h at 4 °C³¹. Subsequently, 110 μl of purified $\Delta 25$ protein at 3 mg ml^{−1} was added to the cholesterol–detergent mixture and incubated for 16 h at 4 °C with gentle stirring. The protein mixture was clarified by two ultracentrifugation steps and used immediately for crystallization experiments. For the $\Delta 25$ -Ca²⁺ structure, the protein was incubated with cholesterol (6 mg) and resuspended in 20 μl of an aqueous solution of 200 mM DDM.

Crystals were obtained at 4 °C using the hanging drop vapour diffusion method. Reservoir solution contained 100 mM Tris pH 8.5–9.5, 150 mM NaCl, 5–20 mM CaCl₂ or BaCl₂, and 29–33% (v/v) PEG 400. Drops were composed of 1:1, 1.5:1, 1.75:1 and 2:1 protein:reservoir, respectively. Crystals typically appeared within two weeks. Crystals were cryoprotected by increasing the PEG 400 concentration in the protein-containing drop to 36% (v/v) before flash cooling in liquid nitrogen.

Structure determination. X-ray diffraction datasets were collected at the Advanced Light Source (ALS) beamline 5.0.2 and at the Advanced Photon Source (APS) beamline 24ID-C and diffraction was measured to $\sim 2.95 \text{ \AA}$ and $\sim 3.2 \text{ \AA}$ for $\Delta 25$ with Ba²⁺ and Ca²⁺, respectively.

Diffraction data were indexed, integrated and scaled using XDS and XSCALE³² software. Diffraction data from $\Delta 25$ -Ca²⁺ crystals were processed using the micro-diffraction assembly method³³. The $\Delta 25$ -Ba²⁺ and $\Delta 25$ -Ca²⁺ structures were solved by molecular replacement using the PHASER program³⁴. For both structures, the extracellular domain coordinates of the Δ ASIC1 structure (PDB code: 2QTS) were used as a search probe. All models were built using iterative rounds of manual model building in Coot³⁵ and refinement in Phenix³⁶ until satisfactory model statistics were achieved. Ramachandran statistics for both $\Delta 25$ -Ba²⁺ and $\Delta 25$ -Ca²⁺ structures were 98.31% favoured and 1.69% allowed, with none disallowed. Omit maps were employed throughout the building and refinement process and to verify the presence of the GAS-domain swap within the second transmembrane domain helix.

Sample preparation, data acquisition, image processing and model building for cryo-EM. ASIC1a was purified as described for $\Delta 25$ with the mobile phase for SEC containing 150 mM NaCl, 20 mM Tris pH 8.0, 1 DDM, 1 mM DTT, 0.2 mM CHS and 5 mM CaCl₂. Peak fractions were concentrated to 3.2 mg ml^{−1} and 2.5 μl of ASIC1a sample was applied to a glow-discharged (15 mA for 60 s on carbon side) Quantifoil Hole Carbon Grid (gold, 1.2 μm hole size, 1.3 μm hole space, 300 mesh), blotted for 3 s at 100% humidity with a VitroBot Mark IV (FEI) and plunge frozen in liquid ethane cooled by liquid nitrogen.

Data were collected on a Titan Krios cryo-electron microscope (FEI) operating at 300 kV. Images were recorded on a Gatan K2 summit direct electron detector, positioned after an energy filter (20-eV slit width), in super-resolution mode with a binned pixel size of 1.04 \AA . Images were collected using the automated image

acquisition software SerialEM³⁷ and dose-fractionated to 100 frames at 0.1 s per frame with a total exposure time and dose of 10 s and 40–50 e[−] \AA^{-2} , respectively. Nominal defocus values ranged from −1 to −3 μm .

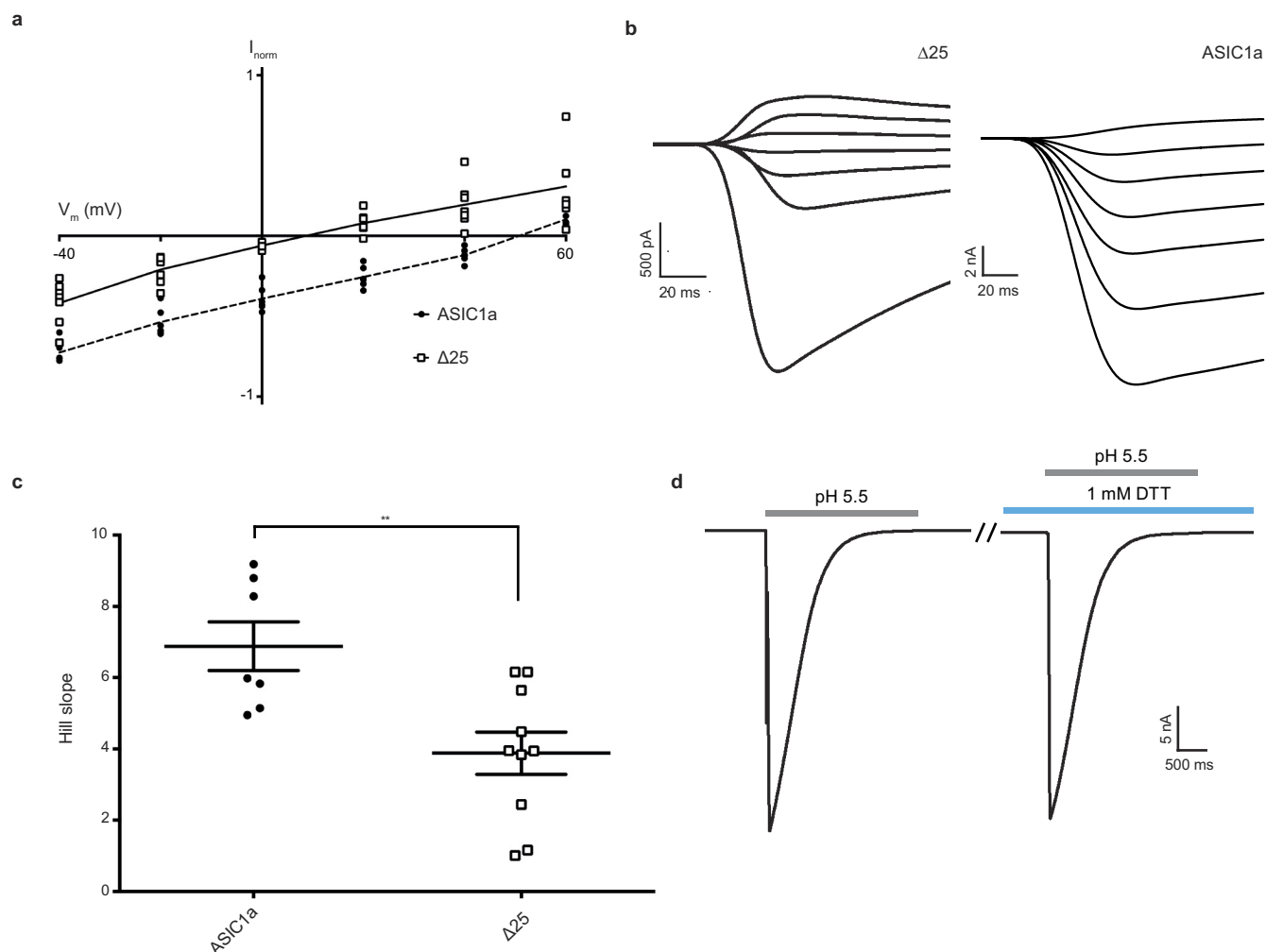
Images were motion-corrected and summed with UCSF MotionCor2³⁸ and defocus values were estimated with Gctf³⁹. Around 256,000 particles were picked using DoGPick⁴⁰ and reference-free 2D classification was performed in Relion⁴¹ to remove broken particles and aggregates. Following 2D classification, $\sim 160,000$ particles were subjected to 3D refinement and classification (C1 symmetry) in Relion to eliminate particles containing some degree of conformational heterogeneity. Inspection of the resulting 3D classes revealed four well-resolved classes containing a total of 33,991 particles, which were subjected to further 3D refinement and classification in Relion (C3 symmetry). Finally, a single class containing 26,117 particles was carried over for final refinement in cisTEM⁴² with C3 symmetry imposed. Refinement in cisTEM was limited to a resolution of 4.5 \AA (≥ 0.9 of the FSC) to prevent overfitting. A mask was used in cisTEM that did not exclude the outlying mask areas, but did filter them to a resolution of 30 \AA to reduce the influence of the micelle on alignment. The final resolution was estimated to be 3.7 \AA based on FSC gold standard analysis in Relion. Local resolution was calculated using blocres from the Bsoft⁴³ package with a box size of 20 and a 0.5 FSC cutoff.

The $\Delta 25$ -Ba²⁺ crystal structure was docked into the EM density map for the ASIC1a channel in Chimera⁴⁴ and served as a template for manual model building of the ASIC1a channel in Coot. Deteriorating density prevented unambiguous modelling of 103 residues corresponding to the channel's cytosolic domains, limiting extension of the model to one and five residues on the amino and carboxy termini, respectively. The final model contains residues 41–464 of chicken ASIC1a and was subjected to real-space refinement in Phenix⁴⁵ and concluded with a CC of 0.867 for all atoms.

Patch-clamp recordings. Whole-cell patch-clamp recordings were carried out on CHO-K1 cells 1–2 days after transfection of plasmid DNA encoding ASIC1a and eGFP separated by an internal ribosome entry site. For characterization of $\Delta 25$ channels and all *I*–*V* experiments, whole-cell patch-clamp recordings were carried out on Sf9 cells 36–48 h after infection with $\Delta 25$ -eGFP or ASIC1a-eGFP P1 BaMam virus⁴⁶. For all electrophysiology experiments, individual cells were used only once for recording; no repeated measurements were taken from the same cell. Pipettes were pulled and polished to 2–4 M Ω resistance and were filled with internal solution containing (in mM): 150 KCl, 2 MgCl₂, 5 EGTA and 10 HEPES pH 7.35. Unless noted, external solution contained (in mM): 150 NaCl, 2 MgCl₂, 2 CaCl₂, 8 Tris and 4 MES. Membrane voltage was clamped at −60 mV. The Axopatch 200B amplifier was used for data acquisition and pClamp 10 software was used for trace analysis.

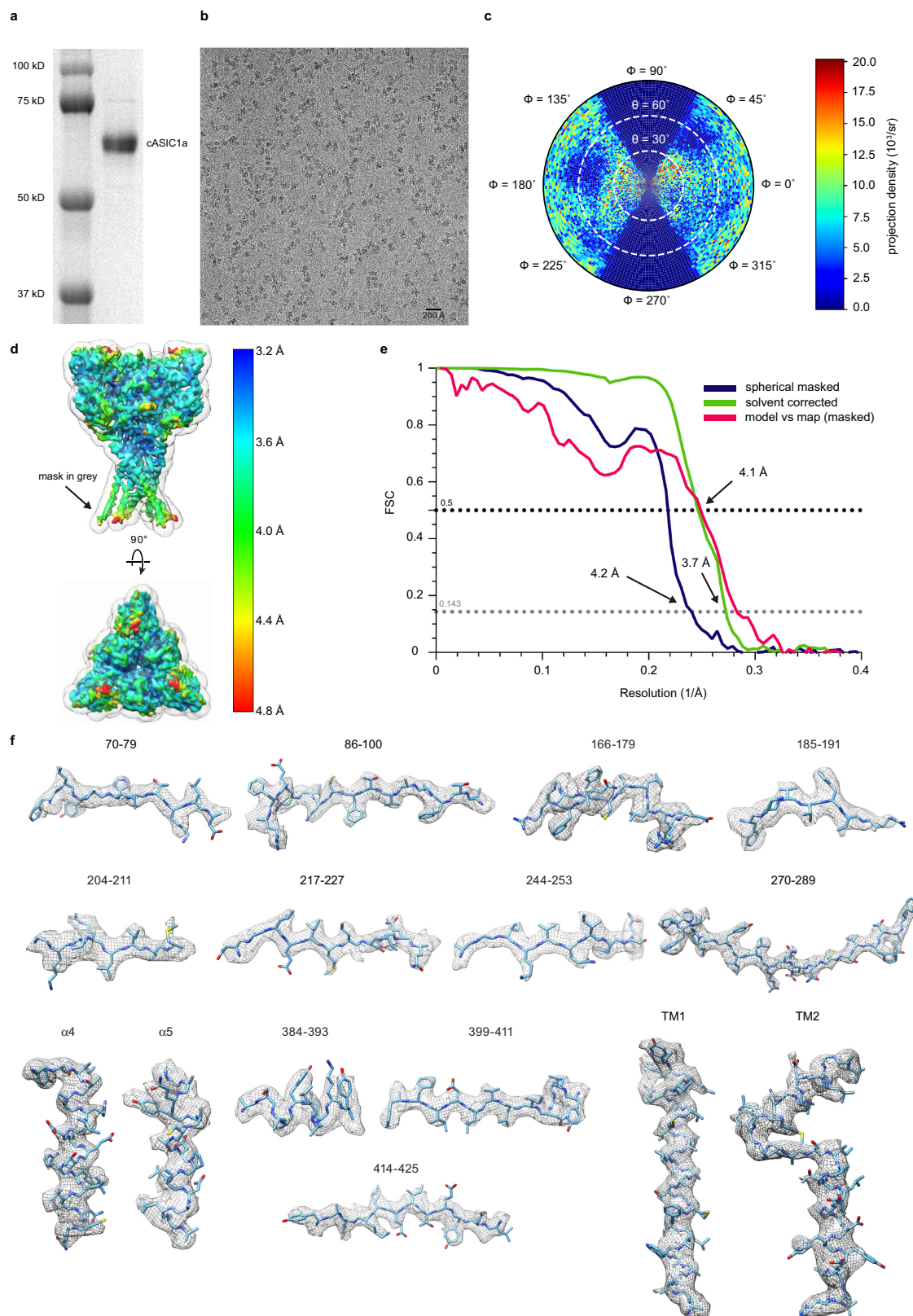
Data availability. The data that support the findings of this study are available from the corresponding author upon reasonable request. The coordinates for the $\Delta 25$ X-ray structures have been deposited in the Protein Data Bank under the accession codes 5WKU and 5WKV. The coordinates and associated volume for the cryo-EM reconstruction of ASIC1a have been deposited in the Protein Data Bank and Electron Microscopy Data Bank under the accession codes 6AVE and 7009, respectively.

- Gourdon, P. et al. HiLiDe—systematic approach to membrane protein crystallization in lipid and detergent. *Cryst. Growth Des.* **11**, 2098–2106 (2011).
- Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
- Hanson, M. A. et al. Crystal structure of a lipid G protein-coupled receptor. *Science* **335**, 851–855 (2012).
- McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D* **63**, 32–41 (2007).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
- Adams, P. D. et al. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
- Mastrorade, D. N. & Serial, E. M. A program for automated tilt series acquisition on Tecnai microscopes using prediction of specimen position. *Microsc. Microanal.* **9**, 1182–1183 (2003).
- Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
- Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
- Voss, N. R., Yoshioka, C. K., Radermacher, M., Potter, C. S. & Carragher, B. DoG Picker and TiltPicker: software tools to facilitate particle selection in single particle electron microscopy. *J. Struct. Biol.* **166**, 205–213 (2009).
- Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
- Grigorieff, N. FREALIGN: An exploratory tool for single-particle cryo-EM. *Methods Enzymol.* **579**, 191–226 (2016).
- Heymann, J. B. Bsoft: image and molecular processing in electron microscopy. *J. Struct. Biol.* **133**, 156–169 (2001).
- Goddard, T. D., Huang, C. C. & Ferrin, T. E. Visualizing density maps with UCSF Chimera. *J. Struct. Biol.* **157**, 281–287 (2007).
- Afonine, P. V., Headd, J. J., Terwilliger, T. C. & Adams, P. D. New tool: phenix.real_space_refine. *Comput. Crystallogr. Newsletter* **4**, 43–44 (2013).
- Goehring, A. et al. Screening and large-scale expression of membrane proteins in mammalian cells for structural studies. *Nat. Protocols* **9**, 2574–2585 (2014).



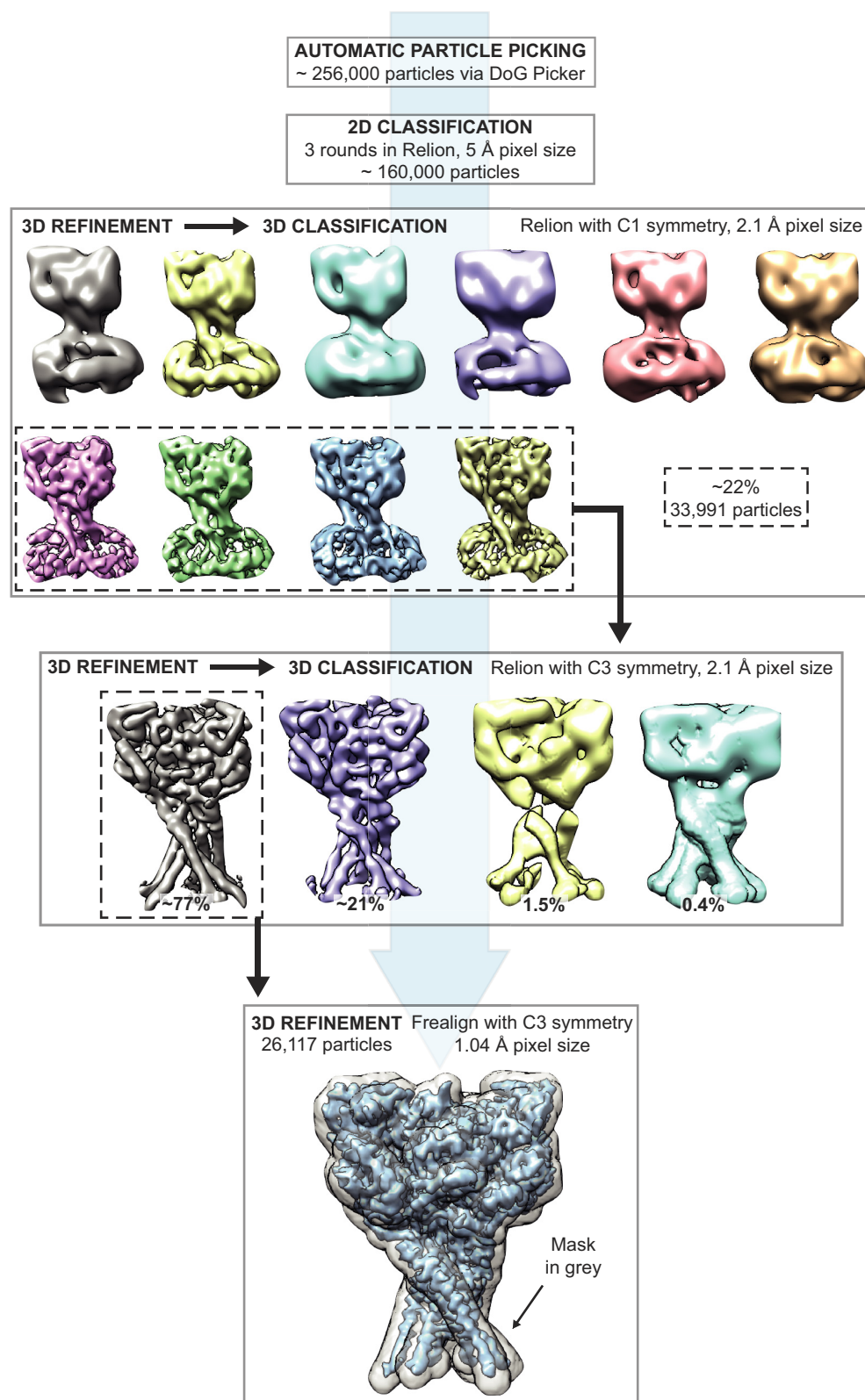
Extended Data Figure 1 | Function of ASIC1a constructs. **a**, I – V relationship for ASIC1a and $\Delta 25$ constructs between -40 and 60 mV. Individual data points are displayed and are normalized to current amplitudes at -60 mV. Lines represent mean. $\Delta 25$: $n = 7$ cells; ASIC1a: $n = 6$ cells. Experiments were performed seven ($\Delta 25$) and six (ASIC1a) times with similar results. **b**, Representative whole-cell patch-clamp recordings at stepped potentials from -60 mV to 60 mV for $\Delta 25$ and ASIC1a. Experiments were performed seven ($\Delta 25$) and six (ASIC1a) times with similar results. **c**, Comparison of Hill slopes of activation for $\Delta 25$ and

ASIC1a channels by unpaired t -test (two-sided, mean \pm s.e.m., $^{**}P \leq 0.01$, $P = 0.0054$; 95% confidence interval = -4.949 to -1.053). ASIC1a: $n = 7$; $\Delta 25$: $n = 10$. Experiments were performed seven (ASIC1a) and ten ($\Delta 25$) times with similar results. **d**, Control experiment demonstrating ASIC1a currents evoked by step to low pH under reducing or ambient conditions. Results are representative of seven independent experiments. Data were collected from Sf9 cells infected with BacMam virus containing $\Delta 25$ or ASIC1a DNA.

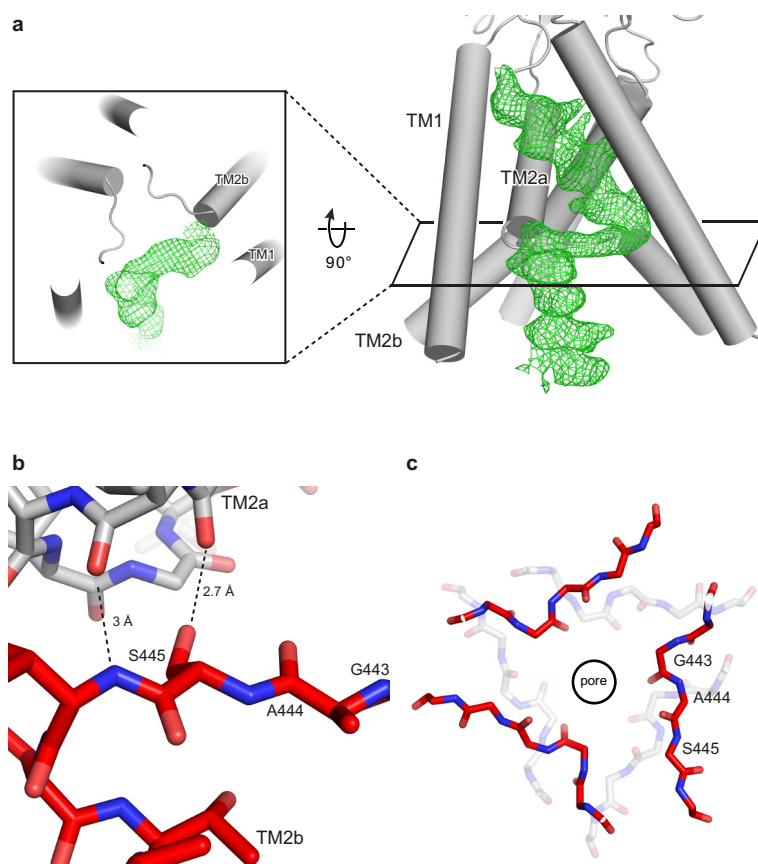


Extended Data Figure 2 | Single particle cryo-EM of chicken ASIC1a.
a, SDS-PAGE of purified chicken ASIC1a. **b**, Representative micrograph of ASIC1a channels embedded in vitreous ice. **c**, Angular distribution of particle projections. **d**, Density map coloured according to local resolution.

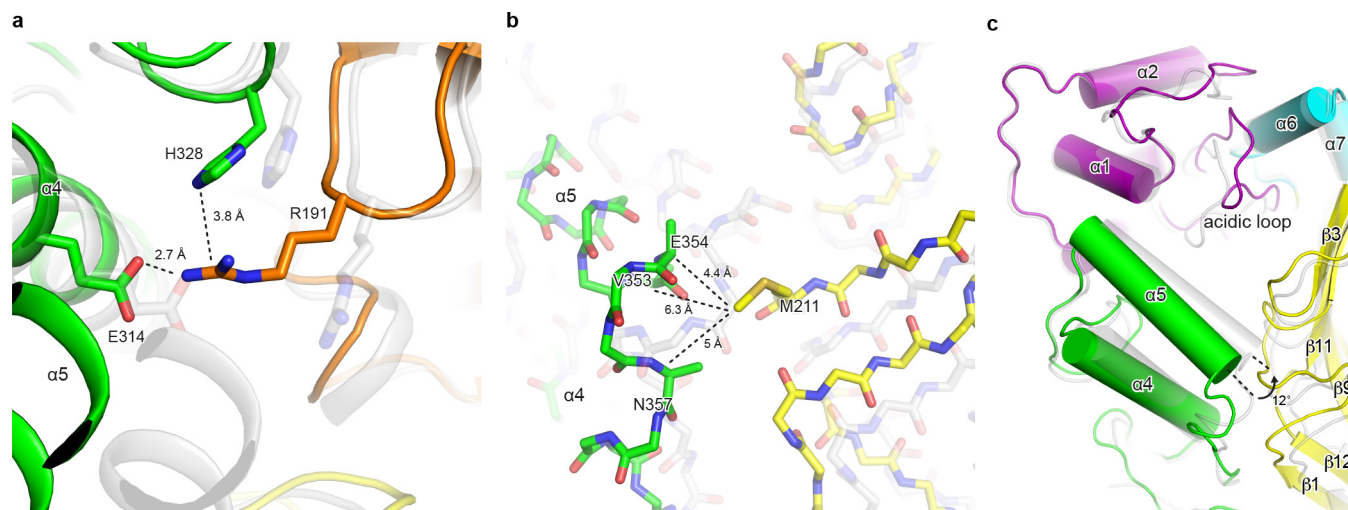
e, Spherical-masked and solvent-corrected FSC curves for density maps and for the refined model to the final 3D reconstruction. **f**, Representative density for the ASIC1a reconstruction, identified by residue range or domain above.



Extended Data Figure 3 | Cryo-EM data processing workflow. Representative data processing steps for the ASIC1a reconstruction.

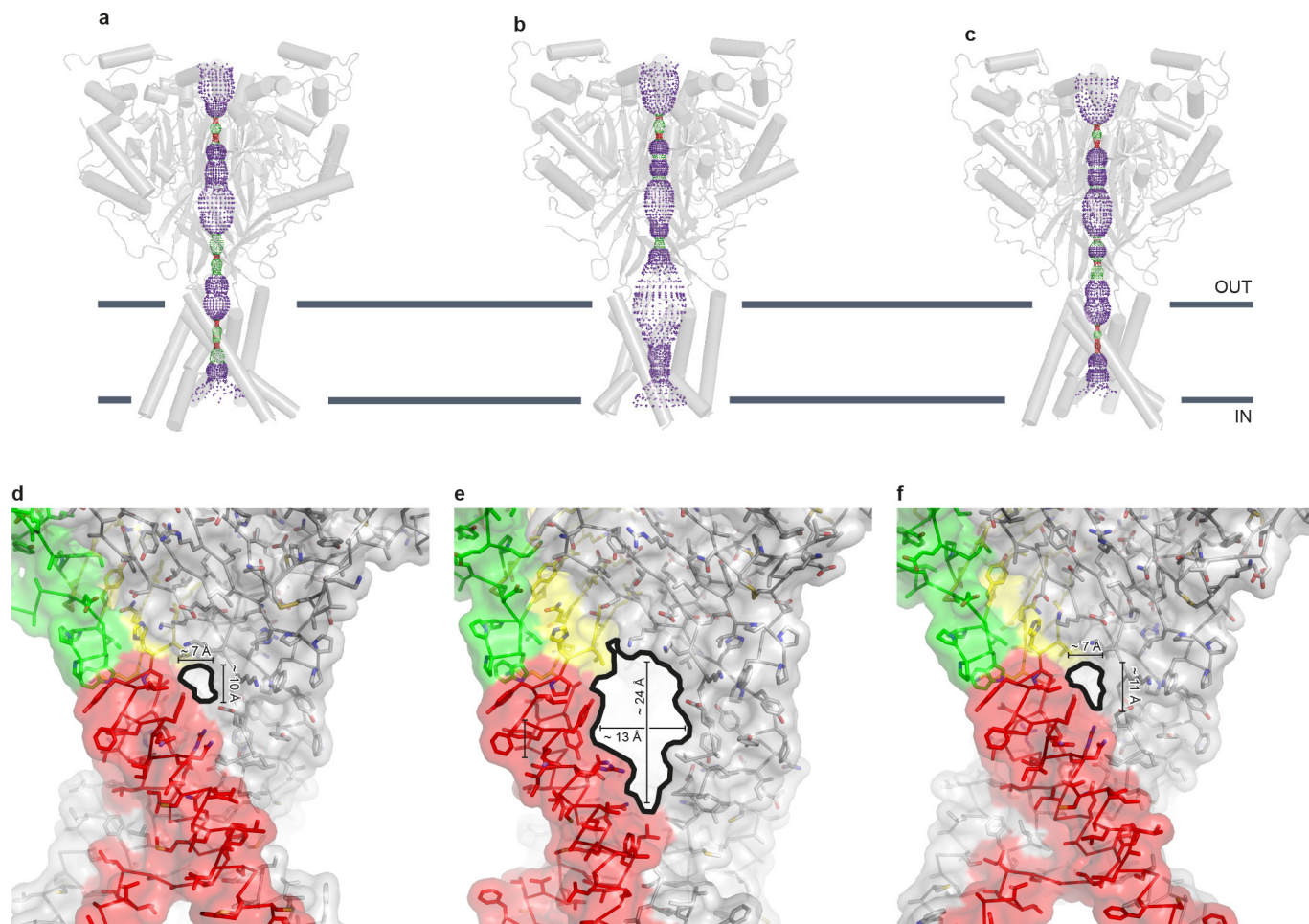


Extended Data Figure 4 | GAS-domain swap. a, Omit map $|F_o| - |F_c|$ density contoured at 2σ for a domain-swapped TM2, top view shown in inset. **b**, Discontinuous TM2 helix stabilized by hydrogen bonds. **c**, Superposition of resting and open (PDB code: 4NTW, grey) channels demonstrates relative conformations of the GAS belt.



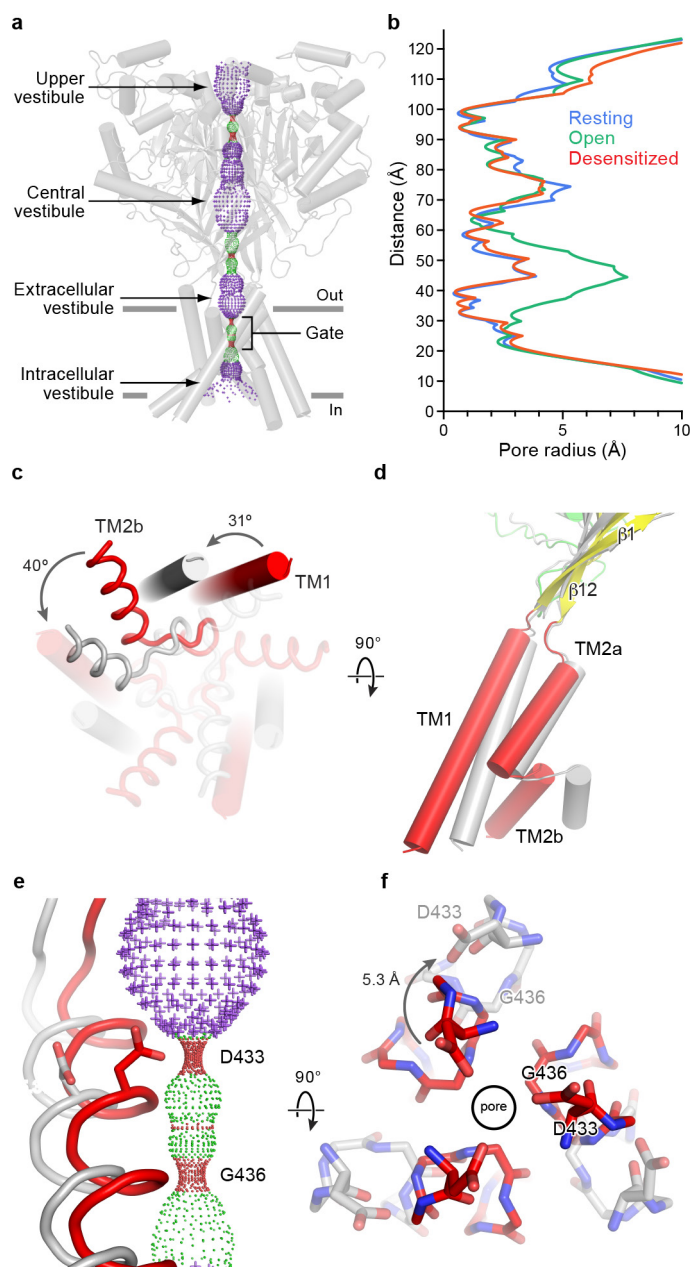
Extended Data Figure 5 | Conformational changes at the acidic pocket.
a, b, Superposition of resting and open (PDB code: 4NTW, grey) channels highlights interactions between Arg191, Glu314 and His328 (**a**) and between Val353, Glu354 and Asn357 with Met211 on an adjacent subunit

(**b**) that stabilize the expanded, high pH conformation of the acidic pocket. **c,** Local superposition ($\alpha 1$ and $\alpha 2$) of resting and open channels demonstrates the $\alpha 5$ pivot upon activation.



Extended Data Figure 6 | State dependence of extracellular fenestrations. a–c, Resting (a), open (b; PDB code: 4NTW) and desensitized (c; PDB code: 4NYK) channel pore profiles calculated with HOLE software (pore radius: red < 1.15 Å < green < 2.3 Å < purple).

d–f, Approximate fenestration sizes for resting (d), open (e) and desensitized (f) channels; approximate fenestration edge is outlined with a solid black line.



Extended Data Figure 7 | State-dependent pore conformation.

a, Resting channel pore profile calculated with HOLE software (pore radius: red < 1.15 Å < green < 2.3 Å < purple). **b**, Plot of pore radius for resting, open (PDB code: 4NTW) and desensitized (PDB code: 4NYK) channels along the three-fold molecular axis. **c**, **d**, Conformation of resting and open TMDs viewed from below (**c**) and the side (**d**). **e**, **f**, Resting and open gates viewed from the side (**e**) and above (**f**).

Extended Data Table 1 | Crystallographic data collection and refinement statistics

	$\Delta 25\text{-Ba}^{2+}$ Resting [†]	$\Delta 25\text{-Ca}^{2+}$ Resting [‡]
Data collection		
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	109.9, 130.4, 157.9	109.2 133.7 157.7
α , β , γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	50 - 2.95	100 - 3.20
<i>R</i> _{meas} (%) [*]	12.4 (549.8)	13.1 (58.1)
<i>I</i> / σ <i>I</i> [*]	15.39 (0.74)	7.05 (2.06)
Completeness (%) [*]	100 (100)	97.5 (99.1)
Redundancy	20.6	5.76
Refinement		
Resolution (Å)	25 - 2.95	25 - 3.2
No. reflections	48366	37646
<i>R</i> _{work} / <i>R</i> _{free}	0.226/0.258	0.287/0.297
No. atoms		
Protein	9247	9263
Ligand/ion	88	120
Water	0	0
<i>B</i> -factors		
Protein	145.62	157.29
Ligand/ion	194.47	187.53
Water	n/a	n/a
R.m.s. deviations		
Bond lengths (Å)	0.004	0.003
Bond angles (°)	0.701	0.688

*Highest resolution shell in parentheses.

[†]Two crystals were merged for the $\Delta 25\text{-Ba}^{2+}$ resting state structure.[‡]Two crystals were merged for the $\Delta 25\text{-Ca}^{2+}$ resting state structure and processed with microdiffraction assembly.5% of reflections were used for calculation of *R*_{free}.

Extended Data Table 2 | Cryo-EM data collection, refinement and validation statistics

	ASIC1a (EMDB-7009) (PDB-6AVE)
Data collection and processing	
Magnification	135000
Voltage (kV)	300
Electron exposure (e-/Å ²)	40 - 50
Defocus range (μm)	-1 to -3
Pixel size (Å)	1.04
Symmetry imposed	C3
Initial particle images (no.)	~ 256000
Final particle images (no.)	26117
Map resolution (Å)	3.7
FSC threshold	0.143
Map resolution range (Å)	3.2 - 4.8
Refinement	
Initial model used (PDB code)	5WKU
Model resolution (Å)	4.1
FSC threshold	0.5
Model resolution range (Å)	
Map sharpening <i>B</i> factor (Å ²)	-100
Model composition	
Non-hydrogen atoms	9657
Protein residues	423
Ligands	6
<i>B</i> factors (Å ²)	
Protein	100
Ligand	100
R.m.s. deviations	
Bond lengths (Å)	0.008
Bond angles (°)	0.946
Validation	
MolProbity score	1.75
Clashscore	6.0
Poor rotamers (%)	0.0
Ramachandran plot	
Favored (%)	95.02
Allowed (%)	4.98
Disallowed (%)	0.0

CORRIGENDUM

doi:10.1038/nature25145

Corrigendum: An immunogenic personal neoantigen vaccine for patients with melanoma

Patrick A. Ott, Zhuting Hu, Derin B. Keskin, Sachet A. Shukla, Jing Sun, David J. Bozym, Wandu Zhang, Adrienne Luoma, Anita Giobbie-Hurder, Lauren Peter, Christina Chen, Oriol Olive, Todd A. Carter, Shuqiang Li, David J. Lieb, Thomas Eisenhaure, Evisa Gjini, Jonathan Stevens, William J. Lane, Indu Javeri, Kaliappanadar Nellaippan, Andres M. Salazar, Heather Daley, Michael Seaman, Elizabeth I. Buchbinder, Charles H. Yoon, Maegan Harden, Niall Lennon, Stacey Gabriel, Scott J. Rodig, Dan H. Barouch, Jon C. Aster, Gad Getz, Kai Wucherpfennig, Donna Neuberg, Jerome Ritz, Eric S. Lander, Edward F. Fritsch, Nir Hacohen & Catherine J. Wu

Nature **547**, 217–221 (2017); doi:10.1038/nature22991

In this Letter, the ‘Data availability’ section in the Methods should state ‘WES and RNA-seq data are deposited in dbGaP (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001451.v1.p1). All other data are available from the corresponding author upon reasonable request.’ instead of ‘All data are available from the corresponding author upon reasonable request’. In addition, the ‘Competing interests’ statement should include: ‘C.J.W. is subject to a conflict of interest management plan for the reported studies because of her competing financial interests in Neon Therapeutics. Under this plan, C.J.W. may not access identifiable human subjects’ data nor otherwise participate directly in the IRB-approved protocol reported herein. C.J.W.’s contributions to the overall program strategy and data analyses occurred on a de-identified basis.’ These errors have been corrected in the online versions of the Letter.

CORRIGENDUM

doi:10.1038/nature25996

Corrigendum: Cholangiocytes act as facultative liver stem cells during impaired hepatocyte regeneration

Alexander Raven, Wei-Yu Lu, Tak Yung Man,
Sofia Ferreira-Gonzalez, Eoghan O'Duibhir,
Benjamin J. Dwyer, John P. Thomson, Richard R. Meehan,
Roman Bogorad, Victor Koteliensky, Yuri Kotelevtsev,
Charles ffrench-Constant, Luke Boulter & Stuart J. Forbes

Nature **547**, 350–354 (2017); doi:10.1038/nature23015

In Extended Data Fig. 10b of this Letter, the axes of the single-cell gating (middle panel) FACS plots were mislabelled. Single cells were gated using forward scatter area (FSC-A) against height (FSC-H) on a linear scale, instead of side scatter area (SSC-A) against height (SSC-H) on a log scale. This does not affect the conclusion drawn. This figure has been corrected in the online versions of the Letter, and the original incorrect figure is provided as Supplementary Information to this Corrigendum, for transparency.

In addition, a reference was inadvertently omitted to earlier work in zebrafish, which should have appeared associated with the sentence 'Biliary cells in zebrafish models have been shown to regenerate the liver after massive hepatocytes loss'. This has been added as ref. 25 in the Letter, and citations in the Methods section (refs 26–28) have been renumbered. The original Letter has been corrected online.

Supplementary Information is available in the online version of this Corrigendum.

25. Choi, T. Y., Ninov, N., Stainier, D. Y. & Shin, D. Extensive conversion of hepatic biliary epithelial cells to hepatocytes after near total loss of hepatocytes in zebrafish. *Gastroenterology* **146**, 776–788 (2014).

CORRIGENDUM

doi:10.1038/nature25998

Corrigendum: Global patterns of declining temperature variability from the Last Glacial Maximum to the Holocene

Kira Rehfeld, Thomas Münch, Sze Ling Ho & Thomas Laepple

Nature **554**, 356–359 (2018); doi:10.1038/nature25454

In this Letter, in the legend of Fig. 3, “Red and green shading” has been corrected to “Green and red shading”. In the Methods subsection ‘Potential effect of ecological adaption and bioturbational mixing on marine variance ratios’, the phrase “alkenone-based $U_{37}^{K'}$ (nine sites) and the Mg/Ca of planktic foraminifera *G. ruber* (six sites)” has been corrected to “alkenone-based $U_{37}^{K'}$ (eight sites) and the Mg/Ca of planktic foraminifera *G. ruber* (seven sites)”. The original Letter has been corrected online.

CAREERS

PERSONAL ETHICS How a vegetarian biologist balances his beliefs with his work **p.405**

BLOG Personal stories and careers counsel
<http://blogs.nature.com/naturejobs>

NATUREJOBS For the latest career listings and advice www.naturejobs.com



DATA MANAGEMENT

For the record

Making project data freely available is vital for open science.

BY QUIRIN SCHIERMEIER

When Marjorie Etique learnt that she had to create a data-management plan for her next research project, she was not sure exactly what to do.

The soil chemist, a postdoc at the Swiss Federal Institute of Technology (ETH) in Zurich, studies the interaction of trace elements in sediments and water. While preparing a grant proposal for the Swiss National Science Foundation last October, she learnt of the funder's new data rules. These require applicants to provide a written plan for the organization and long-term storage of their research data, to help minimize the risk of data

loss and provide guidance for other scientists on how to use the data in the future.

Etique found the task daunting. "Data management is really not my primary skill," she says. "I had absolutely no idea how to go about it." She was able to get advice from her supervisor and from ETH's digital library service. Other researchers might not be so lucky, and might not even know what a data-management plan is — let alone why they would need one and how to produce it. Here, we answer these questions.

WHAT ARE DATA-MANAGEMENT PLANS?

A data-management plan explains how researchers will handle their data during and after a project, and encompasses creating,

sharing and preserving research data of any type, including text, spreadsheets, images, recordings, models, algorithms and software. It does not matter whether the data are generated by large pieces of research equipment, such as imaging tools or particle accelerators, or from straightforward field observation.

Many funders are asking grant applicants to provide data plans. Requirements vary from one discipline to another. But in general, scientists will need to describe — before they begin any research — what data they will generate; how the data will be documented, described, secured and curated; and who will have access to those data after the research is completed. They must also explain any data sharing and reuse restrictions, such as legal and confidentiality issues. Researchers can consult their funder and their host institute's digital library services for assistance. Colleagues who have previously produced data plans may also be able to help (see 'Keeping stock').

WHO NEEDS THEM?

Data management is one example of the way in which public research sponsors and research institutions are implementing 'open science', the push to make scientific research and data freely accessible. Many funding agencies have made data-management plans mandatory for grant applicants in the past decade or so. All US federal agencies, including the National Science Foundation and the National Institutes of Health, have such policies. Data-management plans must also now be included in grant proposals to the European Research Council and other European Union-funded research programmes. And many national funding agencies in Europe — including the UK research councils and the London-based Wellcome Trust, world's largest biomedical research charity — also ask for data plans.

Many scientists already practise data management by default. Astronomers, for example, have been doing so for decades when calibrating their observations and archiving huge amounts of telescope-survey data in standardized, machine-readable catalogues for reuse.

Geneticists, too, use special data repositories to archive the vast amounts of DNA and genome-sequencing data (see go.nature.com/2omlrbe). But less data-intensive fields of science and social research also benefit from data management. For example, geochemists analysing soil bacteria and mineral products in different environments can use it to ►

► collaborate more easily. “In the emerging era of open science, any researcher must be prepared to open up their research processes and results,” says Eloy Rodrigues, library director at the University of Minho in Braga, Portugal, who coordinates FOSTER, an EU-funded open-science e-learning portal.

Still, many scientists are unsure about open-data provisions, and what grant applicants need to do. A 2017 survey of early-career researchers in Europe found that many were unaware of new open-data policies (see go.nature.com/2oqx9s5). Only one-quarter of the 1,277 respondents to the survey, carried out by the European Commission and the European Council of Doctoral Candidates and Junior Researchers (Eurodoc), had actually written a data-management plan; another quarter said they didn’t even know what such a plan might be. Most said they’d not received any relevant training or support from their institutions.

“Data management is inevitably going to be an essential skill in the open-science era,” says Eurodoc’s president, Gareth O’Neill, a linguist at Leiden University in the Netherlands. “And yet, many scientists are scarcely familiar with what it is all about.” The situation in the United States is hardly different, adds Stephanie Simms, a research-data specialist with the California Digital Library (CDL) in Oakland. “We are still at the beginning of a profound shift in research culture,” she says.

WHERE CAN I GET HELP?

The University of California Curation Center, part of the CDL, and the Digital Curation Centre in Edinburgh, UK, provide examples of data-management plans written by researchers from various fields. The centres also provide online tools for writing data-management plans that meet the demands of most funding organizations in both countries. Versions of the tools are also available for scientists in several other European countries, as well as for those in Australia, Canada and South Africa (see go.nature.com/2oquiyz).

Simms recommends that grant applicants who are unfamiliar with open-data provisions consult funding-agency programme officers about any field-specific requirements. For more technical guidance, on requirements for machine readability of data protocols, say, or on file formats used by institutional data repositories, scientists should consult their host institute’s digital library services, she adds.

Etique did just that. Staff members at the ETH’s digital-curation office briefed her about Switzerland’s new open-data policies, and provided her with a generic template for drawing up her data-management plan in line

KEEPING STOCK

Twelve tips for writing a data-management plan

- Check the research-data requirements of your funding agency and field of research.
- Go online for help in developing a data-management plan. A useful guide outlining UK funder expectations can be found at go.nature.com/2tnohla.
- List the various types of data and research outputs that you expect to produce.
- Decide what data and research materials require archiving and determine how much storage space you will need.
- Define appropriate data file formats (see go.nature.com/2tvoo6v for UK formats).
- Look for data repositories used by your research community or your host institution

(see www.re3data.org for examples).

- Check what data format and structure the chosen archive might request.
- Provide metadata that allows others to understand, cite and reuse your data files.
- Make clear how and when your data can be shared with scientists outside your group.
- If your research involves sensitive data, explain any legal and ethical restrictions on data access and reuse.
- Assign responsibility for long-term data curation to a suitable office.
- Revisit your plan frequently and update it if necessary. **U.S.**

with the requirements of the Swiss National Science Foundation.

“It was a bit tricky to address some of the questions, such as file-naming conventions and metadata standards,” she says. But after speaking with information-technology services and ETH library staff, she spent two weeks producing a five-page plan that met all of the funder’s requirements.

Complying with data-management rules is not just another box to tick, says Rachael Ainsworth, an astrophysicist at the University of Manchester, UK. “Your primary collaborator is yourself six months from now, and your past self doesn’t answer e-mails,” says the open-science advocate, who regularly hosts data-management workshops. “So

handling and storing your data in an organized way might save you time and resources.”

DO THE PLANS VARY ACROSS DISCIPLINES?

Data-management demands vary widely, and different research communities (and funders) have different customs and practices. The plans needed for collaborative particle physics, where powerful accelerator facilities generate huge volumes

of experimental data, look very different from those used in smaller research projects, such as Etique’s.

Sarah Jones, a researcher for the Digital Curation Centre who is based at the University of Glasgow, UK, says any data that serve as evidence for a researcher’s claims and results should be archived (the centre was set up in 2005 to champion the management of research data at UK higher-education institutions). This does not mean that a researcher should preserve all of their records, including their lab journal, for posterity, she adds. Indeed, many

scientists whose thesis might rely on a limited number of field observations might need to archive only a small amount of data. And if a project does not generate or reuse any data, as could be the case in purely theoretical science or conceptual work, a data-management plan might not be necessary.

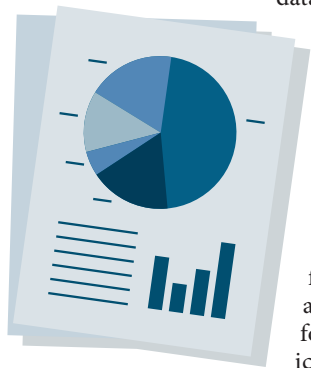
Archived research data must be accompanied by appropriate metadata describing their origin and purpose, so that others will be able to find, read and understand them. Scientists who are unsure about metadata requirements, or about which protocols and digital archives to use for their data proper, should contact their host institute’s library services, says Jones.

Scientists who generate data should specify who will curate the information after the research project is complete. This is essential because scientists spend only so long at a given institute or department. And to guarantee long-term data availability, they should assign that curation responsibility to an office — usually a library department at their current host institute — rather than to a person.

Library departments typically do not curate individual data sets; rather, they archive and maintain institutional repositories so that any data stored there can be accessed indefinitely.

WILL THEY IMPROVE MY SCIENCE?

Access to research data preserves the rights of researchers anywhere to reach independent conclusions about published science. So it’s a good idea for scientists to keep track of their data in case other researchers fail to reproduce the same results, says Jones, or in case legal or ethical problems arise after a paper is published. But not all data types and records can be generously disclosed and freely shared. For example, patient data and health records normally must be anonymized. The same applies to some interview recordings used in empirical social research, such as political surveys or those on personal behaviour.



Data-management plans must also state any constraints regarding confidentiality or copyright, for example. These might relate to collaborations between academic scientists and industry researchers or military services. “Carefully consider data privacy and ethical aspects when writing your plan,” says Ainsworth, adding that ethical, legal or other constraints should be noted.

European research funders will address confusion over open-data policies by setting out minimum standards for discipline-specific data-management plans. The exercise should be completed in a year. “It just doesn’t make sense that different bodies have different rules and requirements when the overarching aims are all the same,” says Peter Doorn, director of data archiving at the Royal Netherlands Academy of Arts and Sciences in Amsterdam, who chairs a joint working group on the topic. “Researchers would rather have clear, not-too-detailed instructions all in one place.”

Scientists needing guidance can check the EU-funded FOSTER portal for webinars and training material on data-management plans (see go.nature.com/2oq4byo). A toolkit, tailored for applicants to the EU’s Horizon 2020 research programme — a 7-year, €77-billion (US\$95-billion) research-funding programme — becomes available in May, says Rodrigues.

Etique, meanwhile, hopes that the data plan that she has submitted with her grant proposal will be reviewed favourably. She expects a funding decision about her project later this year. “It was an opportunity to consider my handling of my research data — it makes sense to think early on about the types and amount of data you will collect with each method and instrument, and how to organize those data for effective use,” she says of her first foray into data management. Such a plan, she notes, can also help scientists to avoid potential problems with data loss and reproducibility. “It may save you a lot of unforeseen trouble,” Etique says.

Unlike the volatile mercury compounds she wants to study, her data are designed to endure. ■ [SEE EDITORIAL P.286](#)

Quirin Schiermeier is Nature’s senior correspondent in Germany.

CORRECTION

The Careers feature ‘Teen spirit in the lab’ (*Nature* **554**, 559–561; 2018) wrongly stated that CERN has 12 member states. In fact, it has 22.

The Careers news story ‘PhD career paths hold promise’ (*Nature* **555**, 277; 2018) gave the wrong year for the start of the PhDs Project. It started in 2000, not 2002.

TURNING POINT

Benefit balancer

Jens Magnusson, a stem-cell biologist at the Karolinska Institute in Stockholm, works extensively with mice, despite becoming a vegetarian in 2012. He explains how he reconciles his personal values with his programme of research.

Why did you become a vegetarian?

It was mainly down to environmental concerns. Meat consumption has a big impact on the environment, partly because animal-derived foods require a lot of energy and resources to produce. I think that as a society, we must shift to a more sustainable diet and that vegetarianism will be a crucial part of the shift. My vegetarianism is motivated by my utilitarian perspective. For me, the positive taste experience that accompanies eating meat is smaller than the negative environmental effect of meat consumption.

Your published research has involved working with animals. Does that pose a problem for you now?

I don’t like the fact that biological research often requires experimentation on animals. However, it is not always possible to replace research animals with other systems such as cell cultures or computer models. For example, when developing pharmaceutical drugs for use in humans, we must find out how the drugs behave in whole organisms before we administer them to people. Similarly, understanding fundamental processes in biology requires the study of such processes in the context of living organisms. I think that the benefits of animal experimentation, which include increasing scientific knowledge and creating new medicines, outweigh the negative ethical consequences such as animal suffering. The advantages of animal research are sufficiently important to motivate experimentation on animals.

Do you have an ethical objection to eating meat or fish?

No. But the more that we learn about how the brains of humans and other organisms work, the more we realize that the subjective experiences of pain, stress and discomfort are not unique to people. I think that this will make it increasingly hard to argue, from an ethical point of view, that animals should be killed for food.

How do you reconcile your concerns with the need to conduct research?

I do not re-evaluate my fundamental moral position every time I do an experiment. Yet,



at the same time, I aim to be guided by my concern for animal welfare. When I plan an experiment, I try to evaluate the level of discomfort that the animals would experience and I let that evaluation strongly guide the experimental design. I also attempt to reduce the total number of animals that will participate in experiments, for example, by using mice that are unsuitable for other researchers’ work.

Do you try to persuade colleagues to become vegetarians?

When I initially became a vegetarian, I was eager to share my thoughts with other people. But I found that this did not really make any of my friends or family change their behaviour.

Has being a vegetarian had an effect on your career progression?

I don’t think it has changed the kind of research projects that I want to be involved in. I feel comfortable with my decision that the benefits of animal research outweigh the moral costs. But I try to re-evaluate that decision occasionally, and if I were to change my mind, I would steer my career towards projects that do not require animal research, such as those involving cell culture.

How do your vegetarian or vegan colleagues balance their own concerns?

Different people have different ways of reconciling their beliefs, some of which seem incoherent to me. For example, one colleague feels guilty about doing research on animals, so they ask another colleague to perform the actual experiments. ■

INTERVIEW BY MARTA PATERLINI

This interview has been edited for length and clarity.